DOCUMENT RESUME

ED 135 821	95	TM 006 044

AUTHOR	Haertel, Geneva D.
TITLE	An Examination of the Relationship Between Memory and Higher Thought Processes Using a Sample of Fifth Grade Children. Technical Report No. 378.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	WRDCCL-TR-378
PUB DATE	Aug 76
CONTRACT	NE-C-00-3-0065
NOTE	102p.
EDRS PRICE	MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS	Achievement Tests; Caucasian Students; Elementary Education; Factor Analysis; Factor Structure; *Grade 5; *Intelligence; Intelligence Tests; *Memory; Recall (Psychological); *Statistical Analysis; *Thought Processes; Validity
IDENTIFIERS	Crystallized Intelligence; Fluid Intelligence; *Jensen (Arthur R)

ABSTRACT

Jensen (1969) proposed a two-factor model of intelligence by which mental abilities can be divided into memory and higher mental processes labeled "Level I" and "Level II" respectively. This study employed statistical methods characteristic of construct validation to examine a battery of memory, intelligence, and achievement tests, seeking evidence of the existence, convergent validity and discriminant validity of the constructs of Level I and Level II intelligence. A principal components factor analysis followed by a promax rotation to oblique simple structure was performed, following Jensen's methodology. To examine the convergent and discriminant validity of each of the constructs, Campbell and Fiske's multitrait-multimethod matrices were constructed. The methods involved were group vs. individual administration for the first matrix and verbal vs. nonverbal response mode for the second. Each of these analyses was performed first using the entire battery of 15 variables. The nine variables used in the second set of analyses were those judged to be "purer" measures of Level I and Level II intelligence. Rural Caucasian fifth grade children were tested. The first factor analysis using all 15 variables yielded six factors providing little support for the two-factor theory. Using all 15 variables there was little support for either the convergent or discriminant validity of the constructs. The factor analysis using the "purer" measures yielded three factors, again providing little support for the theory. Examination of the multitrait-multimethod matrices provided little support for the convergent or discriminant validity of the constructs. In conclusion, the study calls into question the existence of the constructs of Level I and Level II intelligence. In particular, there was little support for the existence of Level I intelligence as a coherent psychological trait. (Author/RC)

TECHNICAL REPORT NO 378

an examination of the relationship between memory and higher thought processes using a sample of fifth grade children

AUGUST 1976

WISCONSIN RESEARCH
AND DEVELOPMENT
CENTER FOR
COGNITIVE LEARNING

Technical Report No. 378

AN EXAMINATION OF THE RELATIONSHIP BETWEEN MEMORY AND HIGHER THOUGHT PROCESSES USING A SAMPLE OF FIFTH GRADE CHILDREN

by

Geneva D. Haertel

Report from the Project on
Conditions of School Learning
and Instructional Strategies

Conrad G. Katzenmeyer
Faculty Associate

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

August 1976

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that agency should be inferred.

Center Contract No. NE-C-00-3-0065

WISCONSIN RESEARCH AND DEVELOPMENT CENTER FOR COGNITIVE LEARNING

MISSION

The mission of the Wisconsin Research and Development Center for Cognitive Learning is to help learners develop as rapidly and effectively as possible their potential as human beings and as contributing members of society. The R&D Center is striving to fulfill this goal by

- conducting research to discover more about how children learn
- developing improved instructional strategies, processes and materials for school administrators, teachers, and children, and
- offering assistance to educators and citizens which will help transfer the outcomes of research and development into practice

PROGRAM

The activities of the Wisconsin R&D Center are organized around one unifying theme, Individually Guided Education.

FUNDING

The Wisconsin R&D Center is supported with funds from the National Institute of Education; the Bureau of Education for the Handicapped, U.S. Office of Education; and the University of Wisconsin.

ACKNOWLEDGEMENTS

This technical report took form over the past two years. Many individuals personally contributed by expressing and discussing with me their beliefs, knowledge, and values. Although I am unable to give personal mention to each one, I wish to name those most involved in its completion.

Dr. Conrad Katzenmeyer willingly provided counsel and support. Mr. Edward Haertel, my husband, applied his skills as a computer programmer and performed the necessary statistical analyses. Finally, a general statement of appreciation goes to the entire R & D Center staff for providing the resources necessary for the completion of this study.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

LIST OF TABLES

LIST OF TABLES (cont.)

ABSTRACT

The relationship of memory to higher thought processes was examined using a sample of rural, Caucasian, fifth grade children.

Jensen (1969), in his article "How Much Can We Boost Scholastic Achievement and IQ?" was the most recent proponent of a two-factor model of intelligence. According to this theory, mental abilities can be divided into memory and higher mental processes labeled "Level I" and "Level II" respectively by Jensen. Considerable research was generated as a result of the theory. Stevenson et al. conducted research studies to examine the performance of lower and middle class preschool children on a sample of tasks chosen as operationalizations of Level I intelligence and Level II intelligence. Tests of predictions derived from Jensen's theory of intelligence provided little support for the theory. The results of Stevenson's study could not be considered to refute the two-factor theory, however, since Level II abilities are not predicted to emerge until after preschool. A more recent study employing multiple operationalizations of each of Level I intelligence and Level II intelligence, and using children at a more appropriate age level, was reported by Jensen in 1973. Jensen factor analyzed a large battery of memory, intelligence, and achievement tests. Three factors emerged, including a factor corresponding to Jensen's Level I intelligence, and two factors which Jensen, using Cattell's terminology, labeled Level II fluid ability and Level II crystallized ability. The present study employed statistical methods characteristic of construct validation to examine a battery of memory, intelligence, and achievement tests, seeking evidence of the existence, convergent validity and discriminant validity of the constructs of Level I and Level II intelligence.

In the present study, a principal components factor analysis followed by a promax rotation to oblique simple structure was performed, following Jensen's methodology. To examine the convergent and discriminant validity of each of the constructs, Campbell and Fiske's multitrait-multimethod matrices were constructed. The methods involved were group vs. individual administration for the first matrix and verbal vs. nonverbal response mode for the second. Each of these analyses was performed first using the entire battery of fifteen variables. The nine variables used in the second set of analyses were those judged to be "purer" measures of Level I and Level II intelligence.

In early fall of 1974, 221 rural Caucasian fifth grade children from a midwest school district were tested, specifically for this study. The first factor analysis using all fifteen variables yielded six factors providing little support for the two-factor theory. Using all fifteen variables there was little support for either the convergent or the discriminant validity of the constructs. The factor analysis using the nine "purer" measures yielded three factors, again providing little support for the theory. Examination of

the multitrait-multimethod matrices provided little support for the convergent or discriminant validity of the constructs. In conclusion, the study calls into question the existence of the constructs of Level I and Level II intelligence as defined by Jensen and earlier investigators. In particular, there is little support for the existence of Level I intelligence as a coherent psychological trait.

The study is significant given the controversy surrounding Jensen's theory of intelligence. Although the study does not examine differences between racial or SES groups, it suggests that Jensen's constructs do not constitute coherent psychological traits, precluding their inclusion in a working theory of intelligence. Furthermore, the results are of use to educators, cautioning them against the use of differential instructional approaches with children from different SES groups, as Jensen suggested.

I

INTRODUCTION

RATIONALE FOR THE STUDY

The relationship of rote memory to higher mental processes has been a subject of speculation since the time of Binet. When Binet formulated his first description of intelligence (Binet & Henri, 1895), he included memory as one of the more complex functions which provided significant information regarding an individual's ability to comprehend. Binet's second formulation of intelligence placed more emphasis on judgment, and he eventually concluded that memory was not among the best predictors of intelligence. By 1905, tests of memory were of secondary importance in the Binet and Simon scale, along with tests of coordination and attention.

Thorndike's major contribution to intelligence theory was contained in the 1927 publication, The Measurement of Intelligence (Thorndike, Bregman, Cobb, & Woodyard, 1927). Thorndike held that intellect was divided into two parts. The lower part of intellect was used in "connection forming," or the association of ideas. The higher part of intellect was involved in abstraction, generalization, and seeing relationships. Within Thorndike's theory, rote memory would require only the lower part of intelligence.

A more recent conclusion concerning the separation of memory and higher mental processes was derived by J. P. Guilford using factor-analytic techniques. Guilford (1956) identified two major classes of mental abilities, which he designated "memory factors" and "thinking factors."

In 1969, A. R. Jensen published a controversial article in the Harvard Educational Review which introduced a theory positing two genotypically based types of intelligence, denoted Level I, "associative ability," and Level II, "conceptual ability" (Jensen, 1969). The theory predicts interactions among types of intelligence, social class, and race.

This report examines the relationship of memory to higher mental processes by the use of various multivariate statistical procedures. Inasmuch as Jensen's formulation of the distinction between memory and higher mental processes is more recent, his terminology has been adopted in the following discussion. The mental ability required for short-term rote memory tasks will be referred to as Level I ability, and conceptual ability will be referred to as Level II. The question of distinctness of memory and higher mental processes can then be phrased as follows: Is Level I ability distinct from Level II ability? This formulation suggests the machinery of construct validation as appropriate for answering the question.

PURPOSE

The purpose of this research was to determine whether the proposed constructs of Level I intelligence (memory) and Level II intelligence (conceptual ability) could be empirically supported when employing several operationalizations of each construct; for this study a sample of approximately 200 white children aged ten to eleven was used.

This research is different from earlier studies in the following ways: (a) There are multiple operationalizations of Level I and Level II intelligence administered to each child; (b) There is a systematic attempt to study the interaction of method of measuring with each of the proposed traits of Level I and Level II intelligence; (c) There is no confounding of race and SES since all children are white, and (d) Children tested were ten and eleven years old allowing for the maximum development of Level I and Level II intelligence.

Several systematic statistical investigations are conducted. Inferential and correlational techniques prescribed by Cronbach and Meehl (1955) and Campbell and Fiske's (1959) multitrait-multimethod matrices are used to examine the adequacy of the Level I and Level II constructs. Operationalizations of the two constructs were chosen so as to allow the construction of two multitrait-multimethod matrices, employing the methods of individual versus group administration and of verbal versus nonverbal response mode. If results had supported the existence of the constructs then additional analyses would have been performed, examining the usefulness of the constructs as well as theoretical predictions concerning social class differences and hierarchical dependence. Inasmuch as the constructs were not substantiated, results of the study were examined with reference to several other theories of intelligence.

RESEARCH HYPOTHESIS

Three major hypotheses concerned with the existence of the constructs of Level I and Level II intelligence as coherent psychological entities are proposed in this study.

Hypothesis I

Factor analysis of a large test battery consisting of multiple operationalizations of both Level I and Level II intelligence will yield three main factors, corresponding to factors which Jensen (1973) labeled fluid intelligence (designated g_f) and crystallized intelligence (designated g_c) with reference to Cattell (1963), both considered types of Level II intelligence, and a factor labeled memory, representing Level I intelligence.

Hypothesis II

Using two multitrait-multimethod matrices as described by Campbell and Fiske (1959), evidence of the convergent validity of each of the constructs of Level I and Level II intelligence will be provided as follows: The intercorrelations among the operationalizations of both of the two constructs in each matrix will be significantly greater than zero.

Hypothesis III

Using two multitrait-multimethod matrices as described by Campbell and Fiske (1959), evidence of the discriminant validity of both of the constructs of Level I and Level II intelligence will be provided as follows: In each matrix, the correlation between any two operationalizations of the same construct will be higher than either of their correlations with any operationalization of the other construct.

II

BACKGROUND AND RELATED LITERATURE

Many symposia held by educators, psychologists, and philosophers have sought to define the concept of intelligence. As early as the fifteenth century men spoke of "intelligentia" as a unitary thing which guided the actions of all men (Spearman, 1927).

In the mid-nineteenth century Herbert Spencer formulated a definition of intelligence which relied upon Darwin's theory of evolution. Spencer, a biologist, defined life as the continual adjustment of internal relations to external conditions. Intelligence was responsible for achieving the adjustment in men, while instinct was responsible in animals. In the light of Darwin's theory of evolution, the concept of intelligence thus became linked with heredity and had obvious implications for the study of individual differences among people (Guilford, 1967, chap. 1).

After Spencer's introduction the concept of intelligence quickly became popular among psychologists. One result of the popularity was the development of numerous intellectual or aptitude tests. The rapid development of the tests, however, was not accompanied by an equally rapid development of curiosity regarding the fundamental nature of human intelligence (Guilford, 1967, chap. 1). Guilford reviews the demands which resulted in the rapid development of these tests. For example, Binet and Simon were commissioned to identify mentally deficient children in French schools. In Germany, mental tests were developed in order to conduct experimental studies in psychopathology and various educational problems. Galtón, in Great Britain, was developing mental tests to study the relationship between heredity and individual differences. Little interest existed in studying the nature of intelligence itself.

With the development of so many mental tests, pressure arose in the psychological community to reach agreement on a definition of intelligence. Numerous symposia were conducted but without consensual agreement on a definition. Spearman (1927) summed up the situation in his assertion that the term "intelligence," with so many diverse meanings, in reality had no meaning at all. A consideration of a few of the diverse definitions of intelligence which were developed will serve to support Spearman's position.

One of the earliest and perhaps most influential members of the psychological community concerned with a practical definition of intelligence was Alfred Binet. His definition of intelligence was never published in a final form (Peterson, 1925), but his point of view has influenced the development of numerous mental tests. Binet did not emphasize sensory and motor functions as did Galton and other British psychologists. Although Binet placed some emphasis upon memory and imagery, he heavily stressed judgment, common sense, and problem solving, all abilities which he believed to involve direction, comprehension, invention and criticism. Recognition of this multitude of abilities caused Binet to regard intelligence as very complex. He viewed mental abilities as being independent and unequal, and expected that, for example, an individual with little memory might be very capable of making judgments. Each of the various subtests he and Simon developed was constructed to measure only one of man's intellectual abilities (Peterson, 1925). Ironically, Binet chose to represent man's intellectual complexity as a single score, the ration IQ. This decision contradicted his belief that intelligence was made up of many abilities. It has been speculated (Guilford, 1967, chap. 1) that the decision for a single score was to facilitate making a single administrative judgment on each child tested.

"Appearing commonly in definitions of intelligence is the statement that it is learning ability or the ability to learn [Guilford, 1967, p. 74]." This conception of intelligence relies upon the assumption that learning ability is a universal trait regardless of the substance or activity that is being learned. Gain scores are used as the index of rate of learning, and it is hypothesized that learning rate is identical to intelligence. Conclusions from experiments using gain scores suggest, however, that "learning ability" is not unitary (Guilford, 1967, chap. 1). Other experiments have attempted to examine the relationship of learning scores to measures of intelligence. In particular, scores on recognized intelligence tests were correlated with gain scores on achievement tests. Guilford also summarized a sample of such studies and concluded that "learning ability" and intelligence involve many different component abilities. The number of components they share depends on the degree of similarity between the learning task (achievement test) and the intelligence measure.

E. G. Boring (1950) directed the efforts of psychologists toward defining intelligence as whatever intelligence tests measure. He suggested the use of this operational definition to eliminate such definitions as "global capability" and "the ability to be rational," which have no empirical referents. Boring made this recommendation in order to decrease the ambiguity of the definition of intelligence used by psychologists as a basis for their test development. One problem with his approach is that each psychologist's freedom to operationalize the construct differently might still result in little information of general significance or usefulness (Guilford, 1967, chap. 1). In light of this, Boring gave some direction for determining a unitary referent for intelligence. He recommended that the method of correlation would provide empirical information to be used in establishing what intelligence tests actually measure. Such empirical investigation is now referred to as construct validation.

The test construction activities of Binet, experimental psychologists' examinations of the relationship of learning ability and intelligence, and Boring's efforts to encourage psychologists to use operational definitions all represented attempts to define intelligence. However, the systematic study of the nature of intelligence to determine its structure as well as its function became the province of primarily two types of theorists: those whose approach is through developmental or genetic investigation and those whose approach is through multivariate analysis.

DEVELOPMENTAL APPROACH TO THE STUDY OF INTELLIGENCE

Jean Piaget provided useful information on the characteristics of the successive stages of intellectual development. His work is considered to be significant since it describes the way particular knowledge develops. In Piaget's conception, intelligence, which is the extension of certain fundamental biological characteristics, can be divided into three aspects: content, function, and structure (Flavell, 1963, chap. 1). Content encompasses the observable behavior of the individual which the psychologist attempts to interpret (Guilford, 1967, chap. 1). Function refers to the principles of intellectual activity common to all ages (Guilford, 1967, chap. 1). Contained within the broad category of function are two principle processes, assimilation and accommodation, which are responsible for the organism's maintenance of equilibrium (Flavell, 1963). Assimilation is the process of taking in information present in the environment. The information becomes part of the individual's existing structure of knowledge through accommodation, which is the process of changing the individual's structure of knowledge in the light of the information being assimilated (Flavell, 1963, chap. 1). Both of these processes are continuous in individuals of all ages. Structure, the third aspect of intelligence, refers to concepts which develop and change with age and experience. Some of the concepts which have been investigated by Piaget include classes, relations, quantity, number, and conservation of quantity and space (Guilford, 1967, chap. 1). As children develop, their use of these concepts becomes more logical in a formal sense (Inhelder & Piaget, 1964). The sequence of development which leads to the use of formal logic is hierarchical in nature, with each stage built upon a previous stage and all of them occurring in an invariant order (Inhelder & Piaget, 1964).

This developmental sequence has several stages. Of particular interest are the stages of concrete operations and formal operations. The concrete stage, typical of children under eight, is characterized by Flavell (1963) as entailing thought bound by events and objects of the real world, and involving little abstraction. Characteristic of adolescents and adults is the stage of formal operations, in which thought is no longer bound by externals but has become more internalized and logical.

Piaget's method is one of clinical observation, not rigorous experimentation. The results of his work have provided psychology with a description of stages of intellectual development through which all individuals appear to pass. Tasks which have been constructed by Piaget to indicate an individual's stage of intellectual development (Guilford, 1967, chap. 7) are used to assess the availability and use of the concepts mentioned earlier.

Piaget has not been interested in the study of individual differences in intelligence per se. Rather, he has been interested in determining and characterizing the stages of intellectual development through which all individuals pass.

MULTIVARIATE APPROACH TO THE STUDY OF INTELLIGENCE

The second approach to studying the fundamental nature of intelligence is represented by those theorists who use multivariate methodology as their primary tool for the investigation of the structure of intelligence. A long series of contributions have been produced by these theorists. In this review the work of each of the following multivariate theorists will be briefly considered: C. Spearman (1927), L. L. Thurstone (1935), C. Burt (1949), P. E. Vernon (1950), R. B. Cattell (1963), J. P. Guilford (1967), and H. E. Garrett (1946).

Spearman (1927) was the first psychologist to use factor analysis in an attempt to explain the structure of intelligence. His work led to the conclusion that intelligence is composed of two basic types of factors: a general factor "G" and a large number of specific factors which he referred to as "s" factors. In his two-factor theory, intelligence is represented primarily by the "G" factor in varying amounts and Spearman (1927) believed it to be the most important aspect of intelligence measured by intelligence tests. "G," or general intelligence, is held to account for the substantial intercorrelations among many intelligence tests, according to Spearman, because all of these tests are measuring this same ability. The most important characteristic of "G" is that it requires insight into relationships.

When Spearman was confronted with evidence of intelligence tests which were not highly correlated, his explanation was in terms of the "s," or specific, factors which were held to be unique to each type of mental ability. An example of a specific or "s" factor was rote memory. In general the "s" factors did not involve insight into relationships.

For Spearman, all mental activities had a common "G" factor which was the most important characteristic. In addition, each mental activity had its own specific ability or abilities. As Spearman continued his work, he recognized that a group of mental tests share some variance in addition to that accounted for by "G." He considered this additional variance to be due to what he called a "group factor." Since Spearman did not view group factors as significant, he did not actively investigate them. Intelligence tests, according to Spearman (1927), should be constructed to measure the "G" factor, and "s" factors should be minimized, thereby leaving "G" as the crucial measure of intelligence.

Spearman's two-factor theory was discounted by Thurstone (1938), who used a different factor--analytic procedure--the centroid method of factor extractions--to examine the structure of intelligence. He noted that often when tests are factor analyzed there is neither a universal "G" factor nor a large number of "s" factors; rather there are several group factors. "Group factor theory," which was Thurstone's contribution to the study of intelligence, is based upon the intercorrelations among limited numbers of tests. He believed it to be the best theoretical base for studying and discussing intelligence (Thurstone, 1938).

Thurstone's original work (1938) included the administration of 56 tests of mental abilities to a large number of adults. Through factor analytic procedures he accounted for performance on the 56 tests with seven basic factors. These seven basic factors represented mental abilities which Thurstone claimed were the primary components of any complex intellectual performance. There was not a high intercorrelation among the seven abilities, so he argued that no general intellectual factor existed. In time, Thurstone's position changed in that he found evidence to support a "second-order factor" which unites the primary group factors (Guilford, 1967, chap. 3).

Followers of Spearman held to his original notation of a "G" factor, but as time progressed and evidence supporting the existence of group factors increased, several of Spearman's followers developed hierarchical models to explain the interrelationships among specific group and general factors (Guilford, 1967, chap. 3). These models depict intelligence as being very specialized. The two major hierarchical models reviewed will be those of Burt (1949) and Vernon (1950).

Burt conceived of a hierarchy of mental abilities representing all of the human mind, with a factor corresponding to each mental ability differentiated at each level of the hierarchy. The human mind contains two general factors, "G" and "practical abilities." Successive dichotomizations occur as the model moves from general factors to group factors, and finally to specific factors. As the model is presented, it departs from strict dichotomizations since there are points at which a higher level ability contains more than two abilities at the next lower level of generality. Consider an example from the Burt model: The entity of the human mind is divided into two abilities, "G" and "practical abilities." "G" contains "memory" and "productive associations." "Memory," in turn, contains several group factors labeled "visual," "auditory," "kinesthetic," and "verbal memory." Each type of memory contains even more specific mental abilities. Throughout the model, abilities at each successive level of generality are labeled: Abilities resulting from the first division, "G" and "practical abilities," are referred to as "relations." Abilities at the second level of generality are called "associations," and third and fourth level abilities are labeled "perceptions" and "sensations" respectively (Guilford, 1967, chap. 3).

The Vernon (1950) model, also hierarchical, postulates "G" as the most important factor. Vernon maintains the existence of major group factors. In turn, each group factor contains more minor group factors, and beyond these factors are very specific factors each having a narrow scope and considered by Vernon to be trivial (Guilford, 1967, chap. 3). For Vernon, "G" is dichotomized into "v:ed" or verbal educational and "k:m" or practical abilities. The "v:ed" factor contains two minor factors: "verbal" and "numerical." The "k:m" factor contains three abilities: "spatial ability," "manual ability," and "mechanical information." Beyond these are the very specific factors.

The concept of "G" as developed by Spearman (1927) has been criticized by several theorists including Thurstone (1938) and Guilford (1967), who have pointed out that its characteristics, rather than being invariant, are entirely dependent on the test battery being used. "G" is not invariant. One of the theorists who moved away from the notion of "G" was R. B. Cattell.

Cattell (1963) developed a theory of intelligence which has two factors: "fluid" and "crystallized." The fluid factor represents an innate ability which requires abstract reasoning but little specific information. Tasks requiring fluid ability necessitate the reorganization of knowledge (Cronbach, 1970, chap. 9). The crystallized factor represents acquired information. Obviously, fluid intelligence remains relatively constant across time with possibly some decline in old age. Crystallized abilities, on the other hand, increase with age.

For Cattell (1963) the notion of fluid ability is intimately associated with culture-fair intelligence tests. Tasks requiring crystallized ability, on the other hand, use information and skills which are the direct result of instruction and are culture-bound.

Guilford (1967), after conducting extensive research, became disenchanted with the hierarchical models which had evolved from earlier theories of intelligence. Research had revealed the existence of many different intellectual factors, and conflicting evidence surrounded "G." Guilford (1967) concluded that even the hierarchical models could not adequately reflect the structure of intellect, and he turned to a morphological model. This model is a cross-classification of intellectual phenomena with intersecting categories rather than nested ones. Guilford's model employs three independent dimensions: content, operation, and product. Any mental ability represents one category in each dimension, resulting in a total of 120 possible intellectual abilities. An intellectual factor results when any one of the five types of operation combines with any of the six types of product and any one of four types of content. The content, operation, and product dimensions are briefly described below.

The content dimension is an extension of the distinction between verbal and nonverbal abilities. The four content categories classify the type of input which is processed. The operation dimension provides five categories of types of processing to be applied to the various types of content. The classification of factors according to product describes the form of the information which is to be processed. Which one of the six product categories is used is determined by the granularity of the input and the types of interrelationships existing in the input.

Each of the theories reviewed represents one of the two systematic approaches to the study of the nature of intelligence. Garrett (1946), however, represented a combination of the genetic and multivariate approaches. He performed factor analyses on scores from a battery of various intellectual tasks administered to children of different ages. Garrett's factor analyses yielded different factors for the various age levels. He concluded that as children mature, mental abilities become more differentiated.

This brief review of the contributions provided by the genetic and multivariate approaches provides some perspective for reviewing the more recent theory of intelligence developed by Arthur Jensen (1969). Jensen's approach to intelligence reflects both the multivariate tradition and the experimental approach, which regards learning ability as equivalent to intelligence.

In developing his model of intelligence Jensen drew heavily from the work of factor analysts for his definition of Level II intelligence, or conceptual ability, which can be viewed as somewhat analogous to Spearman's "G." In developing the argument to support Level I intelligence, or rote memory, Jensen relied upon the tradition of experimental psychology which historically has examined learning ability as a unitary factor independent of content. Jensen contends that the two types of abilities, Level I and Level II intelligence, are hierarchically dependent with Level II processes necessitating the memory abilities which are part of Level I. The notion of hierarchical dependence between abilities was contributed by Burt (1949) and Vernon (1950). White (1965), an American learning theorist who used the experimental method, also influenced Jensen with his theory of hierarchical arrangement of learning processes.

THEORETICAL BASIS FOR THE DISTINCTNESS OF MEMORY FROM CONCEPTUAL THOUGHT PROCESSES

Definitions and Characteristics of Level I and Level II Intelligence

Since Jensen is the most recent theorist to posit a memory factor distinct from higher thought processes, his terminology is adopted in the following discussion. Jensen's (1969) theory posits two types of intelligence. Level I intelligence, originally referred to by Jensen (1969) as "basic learning ability," is the capacity to receive or register stimuli, store them, and later recall or recognize the material accurately (i.e., rote memory). Level I is characterized by "the lack of any need for elaboration, transformation, or manipulation of the input in order to arrive at the output [Jensen, 1970c, p. 4]." The learner need not refer to past learning. Forward digit span is one of the clearest measures of Level I intelligence, whereas reverse digit span is not as pure a measure since transformation of input is required prior to output. Serial rote learning and paired associate learning are also examples of Level I ability. Level I intelligence (Jensen, 1969) is the source of most individual differences on tasks requiring little transformation of input.

Level II intelligence is characterized by transformation and manipulation of the stimulus prior to making the response (Jensen, 1969). Jensen stated that Level II intelligence is the set of mechanisms which make possible generalization beyond primary stimulus generalization, as well as concept formation, encoding and decoding of stimuli in terms of past experience, relating new learning to old learning and transfer of concepts and principles (Jensen, 1969). We shall regard Level II as including all higher mental processes. Spearman's "G," defined as "education of relations and correlates," (Spearman, 1927) would correspond to Level II (Jensen, 1970c). Most standard intelligence tests, especially those such as Raven's CPM which are viewed as culture-fair, would depend heavily upon Level II (Jensen, 1970c). Few tests assess Level I or Level II intelligence in a pure form. Any actual test measures a mixture of these two types of abilities.

Task Sampling as a Method of Establishing Validity of the Constructs of Level I and Level II Intelligence

A research procedure referred to as "task sampling" has been employed as a means of determining whether Level I and Level II intelligence exist. For this purpose, task sampling involves testing the same group of children on several operationally defined measures of Level I and Level II intelligence. Ideally, these measures should be selected such that they elicit either Level I or Level II intelligence exclusively. The tasks may have been used previously in Jensen's studies of Level I and Level II intelligence, or they may be new tasks which meet Jensen's criteria for Level I or Level II tasks.

Stevenson's work is noteworthy in that he uses a number of tasks other than those used by Jensen to test Level I and Level II intelligence. Friedrichs, Hertz, Moynahan, et al. (1971) studied the interrelations and correlates of learning tasks at the preschool level for a group of 50 middle class children. Stevenson, Williams, and Coleman (1971) examined the in rrelations among eight learning and eight performance tasks in four and five year old disadvantaged children. The results of the studies when examined together gave little support to Jensen's position of differences in the intellectual abilities of middle and lower class children. The intercorrelations of the learning tasks were very similar for both social classes. The various operationalizations of Level I intelligence were significantly correlated, and a significant correlation between PA learning and category sorting, representing a significant intercorrelation between Level I and Level II intelligence, was also found. This finding would not be predicted by Jensen (1969) for lower class children. The only performance task, i.e., a task measuring a personality or attitudinal attribute relevant to success in scholastic situations, with a high frequency of significant correlations with the operationalizations of Level I and Level II intelligence was "following instructions." This suggests that the instructions, an aspect of the method of measuring the trait, rather than the content of the tasks themselves, may have introduced variance into the children's performance.

The two studies cited above are valuable contributions to the literature on Jensen's theory. Most importantly, they sample several tasks which ideally should elicit Level I or Level II intelligence. Results of the two studies cannot have maximum impact, however, since at the preschool age Level I and Level II differences between social classes would not yet have become prominent. Jensen (1969) explicitly states that Level II abilities are slow to develop and reach full development between the ages of four and six. SES differences in older children need to be examined using a variety of Level I and Level II tasks.

The high frequency of intercorrelations found between following instructions and learning tasks indicates the need for a systematic study of the type described by Cambell and Fiske (1959) in which the method of measuring a trait is examined through the use of multitrait-multimethod matrices. This method is one of several approaches to construct validation.

THE SYSTEMATIC EXAMINATION OF HYPOTHETICAL CONSTRUCTS

The criticisms of Friedrichs et al. (1971) and Stevenson et al. (1971) raise some doubts as to whether Level I and Level II intelligence exist as coherent psychological traits. Task sampling methods failed to reveal the expected SES differences among preschool children and evidence suggested that the method by which the traits were being measured might be introducing considerable variance into obtained scores (Friedrichs et al., 1971; Stevenson et al., 1971).

Cambell and Fiske (1959) stated that in any attempt to measure a psychological trait, some aspects of the instrument employed are introduced to specifically represent the trait which is to be measured, while other features are intrinsic to the particular method employed. The contribution of method variance limits the validity of obtained scores. Campbell and Fiske's (1959) multitrait-multimethod matrices provided a means of examining the contribution of method variance to test scores. This technique could be employed to systematically examine method variance in various tests designed to elicit Level I and Level II intelligence. Cronbach and Meehl (1955) provided direction for the clarification of constructs by examining patterns of correlation and coefficients.

In defining a construct, Cronbach and Meehl (1955) stated the following characteristics: (a) A construct is a postulated attribute and (b) it is assumed to be reflected in test performance. "A construct has certain associated meanings carried in statements of the general character: Persons who possess this attribute will in situation X act in manner Y (with a stated probability) [Cronbach & Meehl, 1955, p. 284]."

There are several methods which can be used in the investigation of construct validity. A list of methods suggested by Cronbach and Meehl (1955) follows:

1. Use of group differences to test constructs.
2. Correlation matrices and factor analysis.
3. Studies of internal structure (item-test correlations).
4. Studies of process (observation of influences which invalidate the test-taking activity).

Construct validation is the process of confirming or rejecting the tenability of a construct which may not be directly observable by testing the hypotheses it gives rise to. If the hypotheses are confirmed the belief in the construct can be retained. If the hypotheses are not confirmed then there are three alternatives open to the investigator: (1) the operationalizations may be inappropriate, (2) the theory may be inaccurate, or (3) the experimental design may have failed to test the appropriate hypotheses (Cronbach & Meehl, 1955).

An additional method to be used in the study of construct validity was developed by Campbell and Fiske (1959). They assessed convergent and discriminant validity of constructs through the use of correlation matrices, which are referred to as multitrait-multimethod matrices. They were developed to permit the systematic examination of method variance in tests.

It seems that a combination of such construct validity methods could be useful in helping to answer the question of whether the constructs of Level I and Level II intelligence exist as coherent psychological traits.

SUMMARY

Following Spencer's introduction of the concept of intelligence to the psychological community, numerous intelligence tests were developed with little concern for the nature of intelligence itself. To facilitate communication among psychologists, and to permit comparisons of the results of studies using various intelligence tests, pressure arose among psychologists to reach a common definition of the construct "intelligence." Definitions of intelligence were often developed in response to practical needs, but the systematic study of the nature of intelligence was carried out by two major groups of theorists: those employing the genetic approach to intelligence and those relying primarily on multivariate methodologies. Jensen's two-factor theory drew from the multivariate tradition as well as one borrowed from experimental psychology which regards learning ability as equivalent to intelligence.

Jensen (1969) is the most recent proponent of two basic types of intelligence, which he labeled Level I (associative learning ability) and Level II (conceptual learning ability). The terminology of Jensen's theory will be used in this study, which seeks to clarify the relationship of memory to higher mental processes.

Empirical findings on which Jensen's theory is based, as well as subsequent research generated by the theory, have employed essentially four types of statistical methodology: correlation, regression, factor analysis, and analysis of variance. Only a few studies, however, have used measures of Level I and Level II intelligence other than those originally employed by Jensen. Most of these studies used only one measure of each proposed type of intelligence, suggesting the possibility of artifacts arising from "single operationalism (Underwood, 1957)." In many of the studies race and SES were confounded. The results which Stevenson (Friedrichs et al., 1971; Stevenson et al., 1971) obtained when using several new operationalizations of the Level I and Level II constructs were not in agreement with certain predictions of Jensen's theory. Results of these studies did not have maximum impact, however, since the sample used was of preschool age, and according to Jensen's (1969) theory, Level II abilities reach full development between the ages of four and six.

III

METHODOLOGY

SUBJECTS

All fifth grade white children in one Wisconsin public school district whose parents consented to their participation in the research were tested. Of the 273 children in the fifth grade, 221 or 81 percent were permitted by parental consent to participate in the study. A large sample of children were tested in order to avoid attenuation of the range of IQ and scholastic achievement scores, and to assure the stability of the matrix of intercorrelations. Since parental approval for all children to participate was not obtained, the participating children may not be regarded strictly as a representative sample of fifth grade school populations. The treatment of all participants was in accordance with the American Psychological Association's Ethical Principles in the Conduct of Research with Human Participants (1973). No prizes or material awards were provided either during or after the testing.

Several measures of Level I and Level II intelligence were administered to each child in the sample. Table 1 presents the mean and range of IQ using the Peabody Picture Vocabulary Test (PPVT) and achievement scores for the sample. Some of the instruments administered were standardized. Other instruments administered were: measures constructed specifically for use in this research, learning tasks typically employed in laboratory experiments, and standardized instruments for which a nonstandard administration procedure was employed. The results of the pilot tests and detailed descriptions of these instruments are provided below.

In the following sections, each proposed operationalization of Level I and Level II intelligence is presented. Information concerning the reliability of each measure and the rationale for its selection in the light of the following definitions of Level I and Level II intelligence are given.

> Level I ability is essentially the capacity to receive or register stimuli, to store it, and to later recognize or recall the material with a high degree of fidelity. . . . It is characterized especially by the lack of any need for elaboration, transformation, or manipulation of the input in order to arrive at the output. The input need not be referred or related to other past learning in order to issue in effective output. . . . Reception and reproduction of the input with high fidelity is all that is required [Jensen, 1970c, p. 4].

TABLE 1

MEAN AND RANGE OF IQ AND SCHOLASTIC ACHIEVEMENT SCORES FOR THE SAMPLE

Test	Sample $\bar{X}$	Range
Peabody Picture Vocabulary Test IQ	113.70	80-145
Science Research Associates Skills Profile Math Concepts	26.11	7-37
Science Research Associates Skills Profile Math Computation	18.30	3-24
Gates-MacGinitie Vocabulary	28.02	6-41
Gates-MacGinitie Comprehension	31.35	5-50

> Level II ability . . . is characterized by transformation and manipulation of the stimulus prior to making the response. It is the set of mechanisms which make generalization beyond primary stimulus generalization possible. Semantic generalization and concept formation depend upon Level II ability: Encoding and decoding of stimuli in terms of past experience, relating new learning to old learning, transfer in terms of concepts and principles are all examples of Level II [Jensen, 1970c, p. 4].

In addition, information used to classify each measure for inclusion in several multitrait-multimethod matrices is provided. The discussion of these matrices is reserved for Chapter IV.

OPERATIONALIZATIONS OF LEVEL I

Digit Span

The forward portion of the digit span subtest from the Wechsler Intelligence Scale for Children—Revised (WISC-R) was used. This is an individually administered verbal subtest and requires only the registration, storage, and recalling of digits without transformation (Jensen, 1970c). Jensen (1970c) defines digit span forward as a nearly pure measure of Level I. Reverse digit span requires a transformation of input prior to output, and is regarded as a less pure form of Level I ability.

Test-retest reliability of the entire digit span task, as reported in the WISC-R manual (Wechsler, 1974), is approximately .71 for children aged 10 1/2 years. The correlation between the forward portions of the digit span subtests of the Stanford-Binet Intelligence Test and the WISC-R was calculated within each of two levels of SES (Jensen, 1970b). The correlation between the two tests of digit span forward for the low SES group was .49 and for the middle SES group was .62. It should be noted that Jensen administered these two subtests together, interspersing items in order of increasing difficulty. Only two measures entered into these correlations: the longest series a child got correct on the Stanford-Binet Intelligence Test versus the longest series correct on the WISC-R. These correlations represent an attempt to measure the reliability of digit span forward based on a single measure. The reliabilities of the average of the two spans were .65 for low SES and .76 for middle SES. Therefore, digit span tasks appear to be reliable when the dependent variable is the longest series recalled correctly.

Coding B

Coding B is an individually administered nonverbal subtest from the WISC-R requiring a child to match and copy symbols in blank spaces provided on the test sheet using a guide of digits associated with simple shapes. Since children must write the associated symbol, the test uses a free response format. The reason for choosing Coding B and a Level I task is that it is based on the children's learning to combine digits and shapes and then to reproduce these combinations with paper and pencil. There is no transformation of input prior to output. Glasser and Zimmerman (1967) claim that the ability being measured on the coding subtest is the capacity to absorb new material presented in an associative context.

Test-retest reliability of the coding subtest has been established as .76 for children aged 10 1/2 years (Wechsler, 1974).

In this study, the Coding B subtest was administered to classroom-sized groups rather than individuals because tasks were needed which could be classified in the multitrait-multimethod matrices as Level I tests which are group administered. To assess the effects of this nonstandard administration, a pilot test was conducted to establish the correlation between an individual's score when the subtest is group administered and his score when the subtest is administered individually. The verbatim instructions used for the group administration are given in Appendix A.

For the Coding B pilot the design used was a two-by-two factorial model with repeated measures on the second factor. Children were randomly assigned to two groups; one group received first the individual, then the group administration, the second was given the group administration first. The second administration for both groups was given one week after the first administration. The pooled within-group correlation between individual and group administrations was moderately high, $\underline{r} = .77$, $\underline{N} = 31$. A significant practice effect, $\underline{F}_{int}\ (1,29) = 55.426$, $\underline{p} < .001$, may have reduced the magnitude of the correlation coefficient.

Uncategorized Free Recall

Uncategorized free recall is an experimental technique which has been used to test Level I intelligence (Jensen & Frederiksen, 1973). The technique consists of showing each child 20 familiar and unrelated objects one at a time, and after all the objects have been presented asking the child to recall as many of them as he can. Each object is presented for approximately two seconds and the child must name each one when it is presented. If the child is unable to name an object, the examiner will supply a name. The child then has 90 seconds to recall as many of the objects as he can. This procedure is repeated for each of four trials. For each trial, however, the objects are presented in a different predetermined random order. The same four random orders are used for all children. The verbatim instructions to the child are included in Appendix A.

The list of objects used in this task was as follows: ball, bell, book, box, brush, car, cat, coat, couch, egg, flag, frog, gun, key, pen, salt shaker, spool of thread, train, watch, and wheel. These objects were selected to replicate as closely as possible the stimulus materials used by Jensen and Frederiksen (1973). Since both uncategorized and categorized tasks were presented to each child, substitutions were made for 15 percent (three) of the objects to eliminate duplications which occurred in Jensen and Frederiksen's original uncategorized and categorized tasks.

The objects used were not readily grouped into categories, and their recall was more likely to involve rote learning than Level II processes; Jensen and Frederiksen (1973) indicated that it is unlikely that a school-age child would employ Level II processes in performing this task. Uncategorized free recall is classified as a verbal task for the purposes of the multitrait-multimethod matrices.

There is no established reliability on the uncategorized free recall task used by Jensen and Frederiksen (1973). Therefore, a pilot study was conducted to determine the test-retest reliability of the task. Two administrations two weeks apart were given to each of 28 fifth grade children, and test-retest reliabilities were obtained for three variables: the sum of the number of objects recalled on each of the four trials (total amount recalled), the index of subjective organization across trials 3 and 4 ($\underline{SO}_{34}$), and the mean index of subjective organization for trials 1-2, 2-3, and 3-4 ($\underline{SO}_{avg}$) (Tulving, 1962). The reliability of total amount recalled was .41, the reliability of $\underline{SO}_{34}$ was -.10 and the reliability of the $\underline{SO}_{avg}$ measure was -.07.

In Jensen and Frederiksen (1973), the dependent measures used were number of objects recalled for each of the four trials. Total amount recalled is a more appropriate measure for the present study because no differential predictions concerning performance across trials have been made, and sensitive tests of group differences on this single task are not required.

"The SO provides for a measure of sequential redundancy in repeated ordered samples of a set of items. . . . It is simply a measure of actual organization relative to the maximum [Tulving, 1962, pp. 345, 347]." In other words, SO evaluates a child's tendency to recall items in the same order on successive trials, in the absence of any experimentally manipulated sequential organization among the stimulus items (Mandler, 1967).

On the basis of the pilot test results, the only variable used in this study was total amount recalled. The lack of reproducibility in the SO indices supports Jensen and Frederiksen's (1973) classification of uncategorized free recall as a Level I task, although it is possible that stable pairwise clusters would develop if more trials of the task were administered. The poor reliability of total amount recalled in the pilot test may have been due in part to a significant practice effect, $\underline{t}(27) = 7.326$, $\underline{p} < .001$, and to the testing conditions. The categorized and uncategorized free recall tasks were piloted with the same children, in two sittings, with relatively little intervening activity, and some interference was evident to the examiners.

Paired-Associate Task

The paired-associate task (PA) used included eight pictorial PAs. The PAs were presented on 5 x 8 inch index cards. The pictures used, as well as the verbatim instructions read to each child, are included in Appendix A. Each picture is of a familiar object, but the pairs of objects are not typically associated. Each child was tested individually. The task required verbal responses from the children. A practice task consisting of one pairing and one test trial using three sample pairs was administered. The practice test was followed by the presentation of the pairing trial of the 8-item PA list. The pairs of pictures were administered at a rate of one pair of pictures every five seconds. Only one pairing trial was administered, followed by a test trial which was self paced.

A PA task can be defined as Level I if the pairings of elements are not meaningful; in other words, if the elements are not typically associated (Jensen, 1970c). Furthermore, for this task children were not instructed to label the pictures as they were presented, thus further decreasing the likelihood of their engaging in verbal elaboration or other Level II processes.

A pilot study was conducted to determine the test-retest reliability of this task. The task was administered twice to 21 children, with an interval of two weeks between administrations. Based upon experience gained in the course of this pilot the task was shortened from 10 to 8 pairs of pictures. As a result, only seven subjects received exactly the form administered in the major study. The test-retest reliability of the 10-pair task was .54, $N = 21$. The dependent variable used was number of correct responses. A significant practice effect, $t(20) = 3.01$, $p < .005$, may have served to reduce the correlation.

Rohwer, Ammon, Suzuki, and Levin (1970) obtained average test-retest reliabilities of .40, .54, and .56 for kindergarten, first, and third grade children respectively, using parallel forms of a PA task consisting of five pairs of pictures with no verbal labeling. Each of these three correlations is the average of the obtained reliabilities for each of two groups of 48 children, one of low SES black children and one of middle SES white children.

The test-retest reliability for the 10-pair PA task piloted was in the same range as those reported by Rohwer et al. (1970). The decision to use the 9-pair task was made on the basis of the poor performance of fifth grade children on the 10-pair task and the degree of frustration observed in most children during the administration. Over half the children gave only one or two correct responses, and only one child obtained a score of over 50 percent correct.

On the first trial of a 10-pair PA task the mean number of correct responses should be fairly low for the following reason. Miller (1956) suggested that the "magical number" of 7 ± 2 parameterizes the limits of memory. Miller stated the relationship between memory limits and the recoding of information in his utilization hypothesis, which holds that people usually can remember only about seven items from an input list in immediate memory. It is only through recoding or reorganizing items that more information can be remembered. Miller's results suggest that the recall of more than nine items would imply some reorganization of information, i.e., the application of Level II intelligence. Since only one trial was administered in the task under discussion, there is not opportunity for the reorganization of information needed to recall more pairs. Only one trial was administered to obtain as pure a measure of Level I intelligence as possible, since the probability of the reorganization of the information by verbal and imaginal elaboration would increase on successive trials.

The use of the 10-item PA list would have been likely to frustrate children and increase the likelihood of recoding or other Level II processes being employed. The 8-pair task, although piloted on only seven children, appeared to be less frustrating. On the basis of Rohwer's reliabilities on the 5-pair PA task and the obtained results of the pilot of the 10-pair PA task, it was expected that the reliability of the 8-pair task would be comparable.

Visual Memory for Symbols

This nonverbal test, which was group administered, was constructed specifically for use in this research. It made use of symbols similar to those used in tests of visual sequential memory. A placard showing a row of symbols was presented for five seconds; the children were then asked to select the series presented from among four alternates, the three distractors being various arrangements of the same symbols. No verbalization was required from the child.

Seven items were administered, and each of the seven placards was presented for only five seconds. Items increased in difficulty starting with a row of only three symbols and continuing with two items each of six, seven, and eight symbols.

The test was classified as measuring Level I intelligence since it did not require the input to be transformed prior to output. Since Jensen never tested for Level I in the visual modality (Jensen, 1973), this measure gave some indication of the validity of the Level I construct across sensory modalities.

A pilot test was conducted to establish the test-retest reliability of this instrument. The two administrations were conducted over a time span of 18 days. The results of the pilot indicated a fairly low correlation, $r = .48$, $N = 30$. Due to the poor performances on three of the seven items they were dropped from the test, so the test-retest reliability above represents a 4-item test. Various forms of this test were piloted. In each of three pilots, items were dropped due to either a lack of variance or lack of discrimination.

Although the test-retest reliability of .48 would be judged inadequate by many psychometricians, an argument can be made regarding the conditions surrounding a test-retest design and performance when rapid presentation of stimuli is employed. Since stimuli are exposed for only five seconds the opportunity for environmental distractions to interfere with children's performance is great. Since the visual memory task was so sensitive to external conditions during administrations, every effort was made to minimize such distractions during the actual study. However, it was evident to the test administrators that this attempt was largely unsuccessful, and the test results were not used in the analysis of the data. Verbatim instructions used will not be provided here, since results from this test were not included in any analyses.

OPERATIONALIZATIONS OF LEVEL II

All of the operationalizations of Level II intelligence can be classified as either fluid or crystallized intelligence, abbreviated "g_f" and "g_c" respectively, corresponding to Cattell's two aspects of general intelligence (Jensen, 1973). Jensen employed Cattell's definitions of fluid and crystallized abilities in discussing the results of a factor analytic study which he conducted (Jensen, 1973). Following Jensen's usage, the operationalizations of Level II included in the present research are also classified as corresponding to either fluid or crystallized abilities.

OPERATIONALIZATIONS OF LEVEL II FLUID INTELLIGENCE

Fluid intelligence is required for pure abstract reasoning. Fluid tasks require analysis and reasoning, and specific knowledge is seldom helpful. The skills and aptitudes employed in such tasks are not the products of direct training; they involve the reorganization of knowledge rather than the use of skills directly taught (Cronbach, 1970, chap. 9).

Raven's Coloured Progressive Matrices (CPM)

Raven's CPM cannot be solved by simple associative processes and is therefore classified as a Level II test (Jensen, 1969). The solution of the matrices calls for complex inductive reasoning strategies (Jensen, 1970c). The CPM was constructed to load heavily on the "G" factor in the Spearman sense, and on no other ability factors (Jensen, 1969). The test-retest reliability for children nine years of age is reported in the test manual as .80 ± .05 (Raven, 1965). In this study, the test was group administered and for the purposes of the multitrait-multimethod matrices it was classified as a nonverbal instrument.

Block Design

Block Design is an individually administered subtest from the WISC-R. It consists of 11 two-dimensional designs which must be reproduced with multicolored blocks within a time limit. Bonus points are awarded for rapid performance. Kohs (1923) was the first to use copying two-dimensional designs with colored blocks as a nonverbal measure of intelligence. Various descriptions of the abilities and/or processes required for the solution of the block design subtest from the WISC-R support its classification as a measure of Level II intelligence: the task requires analysis, synthesis and reproduction of an abstract pattern, and implicit verbal manipulation (Glasser & Zimmerman, 1967); the ability to reproduce spatial relations involves logic and reasoning (Wechsler, 1949).

The moderately high correlation of block design with the WISC-R full scale IQ, $\underline{r}$ = .65 (Wechsler, 1974) also provides support for its classification as a Level II task. Since block design is a performance task requiring no overt verbalization by the child, it was additionally classified as a nonverbal task for the purposes of a multitrait-multimethod matrix.

The test-retest reliability of the block design subtest for 10 1/2 to 11 1/2 year old children is .85 (Wechsler, 1974).

Categorized Free Recall

The categorized free recall task differs from the uncategorized free recall task only in that a different set of 20 objects is employed. The administration procedure and instructions to children were the same for both free recall tasks; the verbatim instructions are included in Appendix A. The objects used in this task can be grouped into four conceptual categories: animals, clothing, furniture, and tableware. The list of objects by categories follows: chicken, cow, dog, horse, mouse (animals): dress, hat, pants, shoe, skirt (clothing); bed, chair, lamp, sink, table (furniture); and cup, glass, knife, plate, spoon (tableware). These objects were selected to replicate as closely as possible the stimulus materials used by Jensen and Frederiksen (1973). The only substitutions were of pants for coat, since coat was used in the uncategorized list, and of sink for dresser, since no suitable dresser could be found.

Jensen and Frederiksen (1973) state that Level II intelligence may facilitate recall of a set of objects which can be easily categorized, even though it is unlikely that Level II intelligence will be an important determinant of the performance of school age children on the uncategorized free recall task. Since organizing the objects into categories is a conceptual activity, categorized free recall was classified as a Level II task.

There is no established reliability for the categorized free recall task used by Jensen and Frederiksen (1973), so a pilot was conducted to determine the test-retest reliability of the task. Two administrations were given two weeks apart to each of 28 fifth grade children, and the test-retest reliability was obtained for the clustering index, $\underline{Z}$, devised by Frankel and Cole (1971), averaged over the four trials. The reliability of the average of the $\underline{Z}$ index scores over the four trials was .74, $\underline{N}$ = 28.

In Jensen and Frederiksen (1973) the dependent measures used were number of objects recalled and $\underline{Z}$ for each trial. The clustering index should provide a purer measure of Level II intelligence than total amount recalled, as it minimizes the effect of rote memory. The average $\underline{Z}$ index across the four trials was deemed a more appropriate measure for the present study because no differential predictions concerning performance across trials had been made, and sensitive tests of group differences on this single task were not required.

Frankel and Cole's $\underline{Z}$ index is based on the number of runs of items from the same category which occur in a recall list ($\underline{O}_r$). The mean ($\underline{M}_r$) and variance ($\underline{V}_r$) of the number of runs in a randomly selected list of the same category composition as the observed list are computed and a $\underline{Z}$ score is calculated: $\underline{Z} = (\underline{O}_r - \underline{M}_r)/(\underline{V}_r)^{1/2}$. Clustering is defined as the presence of significantly "too few" runs (Frankel & Cole, 1971).

A weakness of the $\underline{Z}$ index is its inability to detect the use of categories other than those the experimenter has established. However, it is superior to Tulving's (1962) index of subjective organization since only a subjective organization measure capable of reflecting bidirectional connections summed over several intervening items would be sensitive to the use of large category clusters (Cole, Frankel, & Sharp, 1971).

The categorized free recall task was classified as an individually administered verbal measure of Level II intelligence.

Similarities

Similarities is a verbal subtest of the WISC-R which estimates the range of an individual's ability to discriminate likenesses or make use of classificatory relationships (Glasser & Zimmerman, 1967). This subtest calls into operation past learning experiences, ability to comprehend, and capacity for associative thinking, as well as the child's ability to select and verbalize appropriate relationships between two dissimilar objects or concepts (Glasser & Zimmerman, 1967). Bialer (1968) states that there is a positive relationship between the ability to acquire secondary generalizations and IQ. Performance on the similarities subtest requires the ability to make such secondary generalizations, which are clearly defined by Jensen (1969) as necessitating Level II intelligence.

Test-retest reliability of the similarities subtest is .81 for children 10 1/2 years old as reported in the WISC-R manual (Wechsler, 1974). For use in the multitrait-multimethod matrices, this task was considered an individually administered verbal test.

OPERATIONALIZATIONS OF LEVEL II CRYSTALLIZED INTELLIGENCE

Crystallized intelligence is required by tests of information or familiar skills. Such tasks require little reasoning; a bright but uninformed person would do poorly on tasks of this nature. Routine application of rules, rapid computation, paragraph reading and table reading are examples of the use of crystallized intelligence. Transfer is required in performing these tasks, but there is little demand for ingenuity (Cronbach, 1970, chap. 9).

Peabody Picture Vocabulary Test

The PPVT consists of 150 sets of four pictures. The examiner says a word and the child must select the picture which best tells the meaning of the word. Items are administered until a basal and ceiling are established. This task was classified for the purposes of the multitrait-multimethod matrices as a verbal test which was given individually.

The PPVT is designed to provide an estimate of a child's verbal intelligence through the measurement of his receptive vocabulary (Dunn, 1965). Verbal intelligence as measured by the PPVT would be considered a Level II process and has been used as such by Rohwer (1970) and Jensen (1970a). Both researchers view the PPVT as more culturally biased than the CPM, but both measures are considered to be operationalizations of Level II ability. The process by which children abstract from a spoken word the salient attributes to be searched for among the pictures requires Level II intelligence.

The test-retest reliability of the PPVT as reported in its manual (Dunn, 1965) was established using alternate forms of the test administered approximately three to seven days apart. For children 10 years old, the test-retest reliability was .77, $N = 319$.

DSU Rhyming Task

This task was suggested by Guilford (1967, chap. 6) as a measure of divergent production of symbolic units (DSU), as described in his "structure of intellect" model. Children are given one minute to write down as many words as possible that rhyme with the word "cat." Verbatim instructions are included in Appendix A. The lists of responses produced are idiosyncratic; no explicit set of rules or processes guide children to correct answers. This task requires Level II intelligence, since transformations of the input word must be made in order to generate a list of rhymes.

A pilot was conducted to determine the reliability of the task. Two group administrations were conducted one week apart, and the test-retest reliability was computed for measures of fluency and originality. The fluency score was simply the number of rhyming words produced. The originality scores were calculated by tabulating occurrences of each rhyming word produced across both administrations. Any word occurring on over 50 percent of the protocols was weighted 1, words occurring on more than 10 percent but less than 50 percent of the protocols were weighted 2, and words occurring on 10 percent or less of the protocols received a weight of 3. Weights assigned to each rhyming word on each protocol were averaged to obtain the originality score for that protocol. This procedure was followed for both the pilot and the main study.

Test-retest reliabilities of .59 for the fluency score and .63 for originality were obtained from the pilot test results ($\underline{N} = 35$). The practice effect was significant for the fluency score, $\underline{t}(34) = 4.61$, $\underline{p} < .001$. For the originality score the practice effect was not significant, $\underline{t}(34) = -.52$.

For use in the multitrait-multimethod matrices, this task was classified as a group-administered verbal measure of Level II intelligence.

Mathematics Portion of the SRA Skills Profile

The computation and concepts subtest raw scores from the mathematics portion of the Science Research Associates (SRA) Skills Profile, battery 2-4, were collected from existing school records with consent of parents. The most recent data available were from the last districtwide test administration during the 1973-74 school year.

The concepts subtest requires children to translate verbal forms into mathematical symbols. Children must also demonstrate their knowledge of the vocabulary of arithmetic as well as their understanding of mathematical principles. The computation subtest requires children to apply the mechanics of computation to problems involving whole numbers, fractions, and decimals (Thorpe, Lefever, & Naslund, 1964).

Coefficients of internal consistency were reported by the test authors. Using the $\underline{KR}_{20}$ formula for each of two parallel forms of the concepts subtest on a sample of 200 fourth grade children for each form, coefficients of .80 and .83 were obtained. $\underline{KR}_{20}$ coefficients for two parallel forms of the computation subtest were .87 and .89 for the same sample. Test-retest reliabilities are not available for these subtests (Thorpe, Lefever, & Naslund, 1964).

Gates-MacGinitie Reading Test

The vocabulary and comprehension subtest raw scores from the Gates-MacGinitie Reading Test, Survey D, Form 2, were collected from existing school records with consent of parents. The most recent data available were from the last districtwide test administration during the 1973-74 school year.

The subtests may be described as follows: The vocabulary subtest samples the child's reading vocabulary. The test contains 50 items. Each test word is followed by five alternatives; the child must choose the word which means most clearly the same as the test word (Gates & MacGinitie, 1965a). Test-retest reliability was reported by the test authors. For the vocabulary subtest the test-retest reliability is .85 (Gates & MacGinitie, 1965b). This reliability was calculated using parallel forms administered less than a week apart to a sample of approximately 780 fourth grade children.

The comprehension subtest measures the child's ability to read prose passages with understanding. There are 21 passages in which a total of 52 blank spaces have been introduced; five alternatives are offered as completions for each blank space. The child must decide which one of the five completions best conforms to the meaning of the whole passage (Gates & MacGinitie, 1965a). The test authors report a test-retest reliability of .83 for the comprehension subtest. This reliability was obtained using parallel forms of the instrument administered less than a week apart to a sample of approximately 780 fourth grade children (Gates & MacGinitie, 1965b).

SUMMARY OF PILOT STUDY RESULTS

The results of pilot studies conducted to establish reliabilities for several of the tests used in this research have been discussed in connection with the respective tests. By way of summary, Table 2 reports the basic statistics and reliabilities obtained in each pilot conducted. As may be seen from the standard deviations, adequate range was obtained on all of the tasks. None of the tasks appears to have been very easy. The PA task was especially difficult, but poor performance is to be expected on the first trial of such a task. Reliabilities reported are not as high as desired, but the strong learning effects in several of the pilots may account in part for this result. Guilford (1967, p. 16) states that "variables that are factor analyzed need not have high reliabilities. Experience shows that a fairly satisfactory factor-analytic solution can be obtained when some variables have as low a reliability as .4." The Level I tasks especially seem less reliable than was expected, but it is argued that with the exception of the visual memory test, which was not used in the analysis of the results of the main study, all are operationalizations actually employed in the study of Level I and Level II intelligence.

PROCEDURE

Two weeks before the main study commenced, a meeting was held with the district's fifth grade teachers. During this meeting the following topics were discussed: purpose of the research, brief review of the tests to be used, confidentiality of results, and logistics of the testing procedures.

After the meeting with teachers, permission slips and an accompanying explanatory letter describing the research and testing situation were mailed to the parents of all fifth graders in the school. A request for family background information and letters of permission to be returned to the school were included. A copy of the letter and parental approval form can be found in Appendix B. Upon receipt of the parental approval forms each fifth grade class in the complying school district was visited and all white children in the fifth grade whose parents consented were included in the study. Each child was tested individually for 50-60 minutes by 1 of 12 trained test administrators. Seven tests were administered during the individual administration. The order in which the tests were presented was the same for all children. The order of tests and approximate times used in administration were as follows:

1. Similarities from the WISC-R (5 minutes).
2. PA Task (3 minutes).
3. Uncategorized Free Recall (11 minutes).

TABLE 2

SUMMARY STATISTICS FOR EACH PILOT STUDY

Test and Variable	Administration 1		Administration 2		Test-Retest Reliability	N	t for Practice Effect
	$\bar{X}$	SD	$\bar{X}$	SD			
Categorized Free Recall Z Index	2.92	.91	3.43	1.26	.74	28	3.20*
DSU Rhyming Fluency	5.69	1.64	6.80	1.49	.59	35	4.61*
DSU Rhyming Originality	1.32	.32	1.30	.19	.63	35	-.52
WISC-R Coding B[a]							
Individual-Group	32.29	9.14	41.64	8.23	.81	14	6.43*
Group-Individual	36.00	7.96	42.59	9.32	.74	17	4.31*
Paired Associate Task	2.81	1.63	4.14	2.39	.55	21	3.01*
Uncategorized Free Recall	51.32	6.11	60.57	6.20	.41	28	7.33*
Visual Memory for Symbols[b]	3.20	.71	3.17	.70	.48	30	-.25

[a] The Coding B pilot was not conducted to assess test-retest reliability, but to establish the equivalence of group and individual (standard) forms of administration. One group received first the group, then the individual forms of the administration, and a second group received the individual administration first. The pooled within-group correlation was .77, N = 31. The correlations for each group are given in the column labeled Test-Retest Reliability.

[b] The pilot results are presented for the final form of this instrument, scored as described in the text. Although this test was administered in the main study, results had to be dropped from the analyses due to poor testing conditions.

*$p < .01$.

4. Digit Span from the WISC-R (3 minutes).
5. Block Design from the WISC-R (10 minutes).
6. PPVT (15 minutes).
7. Categorized Free Recall (11 minutes).

The order of the tests was logically determined to minimize interference among the various subtests. For example, a response set or strategy established during a free recall task could be detrimental to performance on a subsequent PA task. An attempt was also made to maximize the time between the two free recall administrations to reduce interference between the tasks.

All children were also tested in classroom-sized groups for approximately 40-45 minutes by a trained test administrator. Four tests were administered during the group administration. The order in which the tests were presented was determined logically to minimize interference among the tests, and the same order was used for all group administrations. The order in which the tests were administered and the time estimates for each instrument were as follows:

1. Visual Memory Test (5 minutes).
2. Guilford DSU Rhyming (2 minutes).
3. Coding B from the WISC-R (approximate time needed).
4. Raven's CPM (30-35 minutes).

Whenever possible, group administrations occurred after all children in a fifth grade class had received the individual administration. Results of the visual memory test were not used in the main study, due to problems in administering the task.

All 12 test administrators were graduate students from the departments of Educational Psychology, Child Development, and Guidance and Counseling of the University of Wisconsin at Madison. The test administrators were trained for 10 hours regarding the establishment of rapport with examinees, individual test administration procedures, and scoring and statistical properties of tests. Two individual sessions were conducted with each administrator to ensure mastery of the individual administration procedure, and they viewed a demonstration administration of the individual battery to a fifth grade child.

Instructions used for each of the standardized tests or subtests except the WISC-R Coding B subtest were those presented in the administrator's manuals. All test administrators were white, and each tested an approximately equal number of males and females. Test administrators had no information regarding the anticipated outcomes of the study.

Achievement scores for each participating child were obtained from existing school records whenever available. A photocopy of the signed parental approval form giving permission to test the child and enter his/her school records was left in each child's school. Family background information was not photocopied. Achievement data was not collected until all testing in a given school was completed.

Children's names were not recorded on any test or form listing achievement scores. A unique code number was used to identify all the information for any one child, but no association of any child's name with his code number was maintained after data collection in the school was completed. Schools were not provided with individual results, but findings of the research study will be made available upon request.

STATISTICAL ANALYSES

The variables employed in all analyses are specified in Table 3. The two classifications of the variables by methods according to administration type and response mode determine their placement in each of the multitrait-multimethod matrices. For purposes of the first type of matrix, each task in the battery is classified as an operationalization of either Level I or Level II intelligence (trait) employing either a verbal or a nonverbal response mode (method). In the second type of matrix the classification by trait remains the same, but each test is classified by type of administration, either individual or group.

Statistical Analyses Used in Testing the Hypotheses

To test the first hypothesis, i.e., that a factor analysis of all 15 variables would yield three main factors corresponding to two types of Level II intelligence labeled g_f and g_c by Jensen (1973) and a memory factor corresponding to Level I intelligence, a principal components analysis was performed. A promax rotation of the factors with eigenvalues greater than 1.00 was examined following the methodology prescribed by Jensen (1973). It was predicted that in the rotated factor matrix, factors corresponding to memory, g_f and g_c would be identified such that each variable would load .30 or greater on its corresponding factor as identified in Table 3, and less than .30 on other factors. These predictions were made on the basis of Jensen's (1973) findings of a three-factor space of intelligence, which he interpreted with reference to Cattell's constructs of fluid and crystallized intelligence.

A second principal components factor analysis was performed using only 9 of the 15 variables: WISC-R Coding B, PA task, uncategorized free recall, WISC-R Digit Span forward, CPM, WISC-R Similarities, WISC-R Block Design, categorized free recall, and PPVT. The variables not included were: DSU rhyming (neither fluency nor originality scores) and the four scholastic achievement test scores obtained from school records (SRA Math Concepts, SRA Math Comprehension). The reason for performing this second factor analysis was to determine if using a smaller but "purer" set of measures of Level I and Level II intelligence would result in factors corresponding to the constructs of Level I and Level II intelligence. It was not until Jensen's 1973 article and the finding of a three-factor space (g_f, g_c and a memory factor) that Level II was discussed with reference to Cattell's constructs of intelligence.

The DSU rhyming task was dropped from this analysis because it is intended as a measure of creativity; although it requires the ability to transform semantic units, which is a Level II operation, it also involves divergent production (Guilford, 1967), which would not be synonymous with Level II intelligence. Guilford (1950, p. 448) has stated that "we must look well beyond the boundaries of the IQ if we are to fathom the domain of creativity."

The four scholastic achievement tests were dropped from this analysis because they were designed to measure scholastic achievement in specific skill areas. The items included in both the SRA math tests and the Gates-MacGinitie reading tests require the application of school-learned rules and/or procedures. Hence, they are not considered particularly pure measures of Level II intelligence.

To test Hypotheses II and III, the Pearson product-moment correlations and internal consistency or test-retest reliabilities were inserted into four multitrait-multimethod matrices designed as follows. For the first two matrices, all 15 variables were sorted by trait (Level I intelligence versus Level II intelli-

TABLE 3

MULTIPLE CLASSIFICATIONS OF ALL TESTS
BY TRAIT AND METHOD

Test	Method: Response Mode	Method: Administration Type
Operationalizations of Level I Intelligence		
WISC-R Coding B	Nonverbal	Group
Paired-Associate Task	Verbal	Individual
Uncategorized Free Recall	Verbal	Individual
WISC-R Digit Span	Verbal	Individual
Operationalizations of Level II Intelligence (g_f)		
Raven's CPM	Nonverbal	Group
WISC-R Similarities	Verbal	Individual
WISC-R Block Design	Nonverbal	Individual
Categorized Free Recall *Z* Index	Verbal	Individual
Operationalizations of Level II Intelligence (g_c)		
DSU Rhyming Fluency	Verbal	Group
DSU Rhyming Originality	Verbal	Group
PPVT Raw Score	Verbal	Individual
SRA Math Concepts	Verbal	Group
SRA Math Computation	Verbal	Group
Gates-MacGinitie Vocabulary	Verbal	Group
Gates-MacGinitie Comprehension	Verbal	Group

gence) within method, as indicated in Table 3. For the first matrix the method used to group the variables was response mode (verbal versus nonverbal). For the second matrix, the method used to group the variables was administration type (group versus individual). The third and fourth matrices were formed using exactly the same trait and method categories, but included only the 9 variables which were used in the second factor analysis, described above as "purer" measures of Level I and Level II.

Hypothesis II, regarding the convergent validity of the constructs, was tested by examining the magnitudes of the correlations within the multitrait-multimethod matrices. It was predicted that each correlation between two operationalizations of the same trait (Level I intelligence or Level II intelligence) would be significantly greater than zero. This hypothesis was tested by the use of Fisher's r to Z transformation.

Hypothesis III, concerning the discriminant validity of the constructs, stated that the correlation between any two operationalizations of the same construct will be higher than either of their correlations with any operationalization of the other construct. Since only relative magnitudes of intercorrelations are examined, no tests of significance were performed. Campbell and Fiske (1959) examined only the relative magnitudes of correlations in their discussion of discriminant validity as evidenced in the multitrait-multimethod matrix. They do not recommend the use of statistical tests for examining discriminant validity, but do suggest significance tests when examining convergent validity.

Supplementary Statistical Analysis

One additional statistical investigation was performed: the Pearson product-moment correlations of scholastic achievement as measured by the SRA Mathematics Achievement Test and the Gates-MacGinitie Reading Test with each of the other variables were computed. Tests of significance using Fisher's r to Z transformation were performed to determine whether each correlation was different from zero.

LIMITATIONS OF THE RESEARCH

To facilitate appropriate interpretation of the results which are presented in Chapter IV, certain methodological and procedural weaknesses of the study will be briefly indicated.

While test administrators were carefully trained, the conditions under which individual administrations were conducted varied widely from school to school. In several schools, it was impossible to secure facilities for quiet and/or uninterrupted individual testing. The collection of data under adverse conditions may have reduced the magnitude of the intercorrelations among the various measures.

The process of choosing tests to serve as adequate operationalizations of Level I and Level II intelligence requires adequate definitions of the constructs. Jensen's definitions specify internal processes, with little discussion of observable behavior. Therefore it is difficult to assure the adequacy of the operationalizations employed. In particular, there is some evidence that some operationalizations of Level I elicited Level II processes in at least some of the children tested.

The correlations of the PA task and of the CPM with the other measures may have been attenuated due to the levels of difficulty of these two tasks. Performance on the PA task was generally poor, and not much variation was obtained. On the CPM, some ceiling effect was evident.

IV

RESULTS

Following a brief presentation of descriptive statistics on each variable and the correlation of sex with each variable, the results of the study will be presented with regard to the three hypotheses tested; a supplementary analysis relating scholastic achievement to each of the other variables will also be given. The section on the first hypothesis will include the results of the principal components analysis followed by a promax rotation using all 15 variables, as well as the results of the same statistical procedure using a subset of 9 variables which are "purer" measures of the constructs. T. section on the second hypothesis will include presentation of the four multitrait-multimethod matrices and the results of the significance tests concerning the convergent validity of the constructs. The section concerning Hypothesis III presents tabulations of the relative magnitudes of various pairs of correlations, as described in the statistical procedures section of Chapter III. The fourth section will present the results of the supplementary statistical investigation regarding the correlation of scholastic achievement with each of the other variables.

DESCRIPTIVE STATISTICS

The mean, standard deviation, minimum, and maximum for each of the 15 variables used in the study are presented in Table 4. In general the tests appeared to be of appropriate difficulty, although the mean of the PA test is quite low. This was to be expected with only one study trial being presented. The range of scores on the PA test indicates that there were children capable of attaining the maximum score. Examination of the CPM results indicates the possibility of a ceiling effect, but the test is reported to be appropriate for use with children through age 11 (Raven, 1965). The mean age of this sample was 10 years 7 months; no children were included in the data analysis who had reached age 12.

The point-biserial correlation of sex with each of the variables is given in Table 5. Of the 15 point-biserial correlations reported, 6 were significantly different from zero at an alpha level of $\underline{p} < .05$, and 4 of those were significant at an alpha level of $\underline{p} < .01$. These correlations were computed on the entire sample of 221 children.

For the two free recall tests, the stimulus materials employed included doll house furniture and doll clothes; these may have been of more intrinsic interest to girls than to boys, contributing to the significant correlation of sex with these two tests. No explanation is offered for the remaining signi-

TABLE 4

MEAN, STANDARD DEVIATION, MINIMUM, AND MAXIMUM FOR EACH VARIABLE

Variable	Maximum Possible Score	$\bar{X}$	SD	Minimum	Maximum
WISC-R Coding B[a]	100	41.90	7.56	22	62
Paired-Associate Task[a]	8	1.54	1.37	0	8
Uncategorized Free[a] Recall	80	49.92	6.17	29	68
WISC-R Digit Span[a]	14	6.21	1.81	2	13
Raven's CPM[a]	36	30.66	3.60	18	36
WISC-R Similarities[a]	30	15.00	3.43	7	24
WISC-R Block Design[a]	62	29.09	10.66	2	54
Categorized Free[a] Recall Z Index	6.75	2.51	.96	-.63	4.83
DSU Rhyming Fluency	none	5.84	1.50	1	10
DSU Rhyming Originality	3	1.32	.21	1.00	2.00
PPVT Raw Score[a]	150	88.76	7.90	69	112
SRA Math Concepts	47	26.11	5.54	7	37
SRA Math Computation	47	18.30	4.78	3	24
Gates-MacGinitie Vocabulary	50	28.02	7.32	6	41
Gates-MacGinitie Comprehension	52	31.35	10.07	5	50

[a]These variables comprised the nine "purer" measures used in the second factor analysis and the third and fourth multitrait-multimethod matrices.

TABLE 5

POINT-BISERIAL CORRELATION OF SEX WITH EACH OF THE FIFTEEN VARIABLES INCLUDED IN THE STUDY

Variable	Correlation
WISC-R Coding B	.293**
Paired-Associate Task	-.104
Uncategorized Free Recall	.193**
WISC-R Digit Span Forward	-.054
Raven's CPM	-.020
WISC-R Similarities	-.140*
WISC-R Block Design	-.156*
Categorized Free Recall-Z	.239**
DSU Rhyming Task-Fluency	.074
DSU Rhyming Task-Originality	.034
PPVT Raw Score	-.222**
SRA Math Concepts	-.049
SRA Math Computation	.114
Gages-MacGinitie Vocabulary	-.060
Gates-MacGinitie Comprehension	.016

*$p < .05$

**$p < .01$

ficant correlations. Since it was not the intention of this study to examine sex differences, separate analyses of the data for boys and for girls were not performed.

HYPOTHESIS I

Hypothesis I states that the factor analysis of a large test battery consisting of multiple operationalizations of each of Level I and Level II intelligence will yield three main factors corresponding to the two types of Level II intelligence delineated by Jensen (1973) with reference to Cattell's (1963) theory, and a memory factor representing Level I intelligence.

In order to test this hypothesis all 15 test scores for each child were intercorrelated, chronological age was partialed out of all intercorrelations following the methodology used by Jensen (1973), and the resulting matrix was subjected to a principal components analysis. FACTOR3, a computer program maintained by the University of Wisconsin Academic Computing Center, was used to derive the principal components (Academic Computing Center, 1974). Factors corresponding to eigenvalues greater than 1.00 were then rotated to oblique simple structure using the promax procedure described by Hendrickson and White (1964). Six factors emerged with eigenvalues greater than 1.00. The eigenvalues corresponding to the six unrotated factors, which accounted for 65.4 percent of the total variance, are 3.7723, 1.5563, 1.2142, 1.1593, 1.0630, and 1.0385. Table 6 presents the oblique factor loadings. The intercorrelations among the factors, presented in Table 7, were all less than .31 in absolute value, suggesting that the six factors are relatively independent of one another.

Significant factor loadings are summarized in Table 8. The only variables which load significantly on the first factor are the SRA math concepts and computation subtest scores, suggesting that the first factor might be a school-related mathematical ability factor. Four of the 15 variables load significantly on the second factor: WISC-R similarities, PPVT raw score, and both the vocabulary and comprehension subtest scores of the Gates-MacGinitie Reading Test. These four measures support the interpretation of the second factor as a language ability factor heavily saturated with size of vocabulary. The third factor is not as school related as the first two factors discussed. The variables loading significantly on the third factor are Raven's CPM and the WISC-R Block Design subtest, both tests of skills which are not taught as part of the typical school curriculum. Neither of these tests require overt verbalizations although both may depend upon covert verbalizations. The cognitive operations required for successful performance on both of these tests have been described by various psychologists as requiring complex reasoning strategies (Wechsler, 1949; Glasser & Zimmerman, 1967; Raven, 1965). Since both of the tests have been characterized as requiring complex reasoning strategies which may in turn require implicit verbal manipulations in order to either reproduce or complete a matrix of visual stimuli, it is insufficient to refer to the third factor as a perceptual or spatial ability factor. Rather, the factor might be labeled as abstract reasoning using figural content.

The PA test and the digit span test are the 2 variables which load significantly on the fourth factor; clearly it can be labeled a short-term rote memory factor. This interpretation can be supported since there was only one study trial for the PA test and the use of verbal labels or production of sentences or stories to facilitate memorization of the two pictures comprising a pair was not encouraged. Therefore, successful performance on both digit

TABLE 6

FACTOR STRUCTURE OF THE ROTATED PRINCIPAL COMPONENTS OF THE FIFTEEN VARIABLES

Variable	Component					
	I	II	III	IV	V	VI
WISC-R Coding B	.020	-.272	.089	.308	-.015	.466*
Paired-Associate Task	-.204	-.404	.105	.729*	-.127	-.021
Uncategorized Free Recall	.276	-.089	.211	-.106	-.140	.287
WISC-R Digit Span	.011	.041	-.106	.609*	.026	-.288
Raven's CPM	.167	.059	.599*	.048	.046	-.052
WISC-R Similarities	-.114	.669*	.291	.072	-.199	.013
WISC-R Block Design	.069	.095	.804*	.005	.132	.090
Categorized Free Recall Z Index	-.192	.145	-.140	-.201	.011	.852*
DSU Rhyming Fluency	.029	.019	.125	.090	.619*	.103
DSU Rhyming Originality	-.083	-.081	.102	-.137	.854*	-.043
PPVT Raw Score	-.051	.805*	.058	-.118	-.006	.012
SRA Math Concepts	.727*	.127	.008	-.183	.001	-.102
SRA Math Computation	.774*	-.159	.020	-.052	-.092	-.138
Gates-MacGinitie Vocabulary	.253	.456*	-.160	.251	.058	.059
Gates-MacGinitie Comprehension	.281	.410*	-.124	.129	.091	.155

*Factor loadings significant at the .01 level (Harmon, 1967).

TABLE 7

INTERCORRELATIONS AMONG THE SIX OBLIQUE FACTORS

Factor	I	II	III	IV	V	VI
I	1.000					
II	-.299	1.000				
III	-.208	.068	1.000			
IV	-.288	-.133	.005	1.000		
V	-.154	-.118	.094	-.114	1.000	
VI	-.307	.110	-.227	-.148	.005	1.000

TABLE 8

SIGNIFICANT LOADINGS ON THE ROTATED FACTORS FROM
THE ANALYSIS OF ALL FIFTEEN VARIABLES

	Factor					
Variable	Mathematical Ability	Language Ability	Abstract Reasoning Using Figural Content	Short Term Rote Memory	Creativity	Mnemonic Facility
WISC-R Coding B						*
Paired-Associate Task				*		
Uncategorized Free Recall						
WISC-R Digit Span				*		
Raven's CPM			*			
WISC-R Similarities		*				
WISC-R Block Design			*			
Categorized Free Recall Z Index						*
DSU Rhyming Fluency					*	
DSU Rhyming Originality					*	
PPVT Raw Score		*				
SRA Math Concepts	*					
SRA Math Computation	*					
Gates-MacGinitie Vocabulary		*				
Gates-MacGinitie Comprehension		*				

span and the PA test required immediate recall of a set of stimuli which were purposely selected to discourage any verbal mediation. This factor is very close to Jensen's Level I factor. The two variables which load significantly on the fifth factor are the DSU fluency and originality scores. Since both of these scores loaded together on this factor, it seems appropriate to label it a creativity factor.

The WISC-R Coding B subtest and the categorized free recall Z index load significantly on the sixth factor. Although some of the characteristics of these two tests seem quite different, they are alike in that they provide an opportunity for mnemonic strategies to develop. In the coding test, the child is presented with a key of numbers and arbitrary symbols such that each number is associated with a unique symbol. As described in Chapter III, the child is given two minutes to fill in a grid where a row of numbers is presented; s/he must fill in the box below each number with the appropriate symbol. The key with numbers and symbols is present before the child the entire time during which s/he is performing, thereby providing an opportunity to develop a facilitative mnemonic strategy. An example of such a strategy might be giving one of the symbols a label, for example, calling the symbol "⊣" a "sideways T." Such labels may help the child recall the way the symbol looks and facilitate production of the grapheme. In the categorized free recall test, the child is presented with the same 20 objects in succession on each of four trials. On each trial the 20 objects are presented in a different random order. The child has ample opportunity to become aware of the a priori categories implicit in the stimulus objects (five animals, five pieces of clothing, five pieces of furniture, five pieces of tableware).

Bousfield and his associates demonstrated that subjects will recall randomly presented materials in clusters (Bousfield, 1953). It was also demonstrated that as the number of presentations of the stimulus increases, the clustering tendency increases (Bousfield & Cohen, 1953). More importantly, there is a strong correlation between the degree of clustering and the amount of recall (Bousfield & Cohen, 1955). The fact that there are four trials during which these categories may come to be recognized encourages the development of mnemonic strategies. The variable being analyzed for the categorized free recall test is the mean value of the clustering index, Z, across the four trials. This index is a measure of the amount of clustering evident in the set of objects recalled. It reflects the degree to which the particular mnemonic strategy of clustering the objects with respect to the experimenter-provided categories was employed. In other words, on both the coding and categorized free recall tests the child is repeatedly exposed to the stimuli, allowing for the development of useful mnemonic strategies. In further support of this interpretation, the variable with the next highest loading on this factor was the free recall uncategorized test. Given that there are repeated presentations of the stimuli in this test also, performance could again be facilitated through the use of a mnemonic strategy. In view of this interpretation, the sixth factor is referred to as a mnemonic facility factor.

The obtained six-factor solution bears little resemblance to the three-factor solution predicted in Hypothesis I. The four variables which were chosen as operationalizations of g_f or Level II fluid intelligence should have loaded on a single factor; instead they loaded significantly on three of the six factors. To summarize, the placement of the four variables predicted to load on the g_f factor is as follows: the WISC-R similarities subtest loaded heavily on the language ability factor, Raven's CPM and the WISC-R block design subtest loaded on the abstract reasoning using figural content factor, and the categorized free recall test loaded on the mnemonic facility factor. The six variables predicted to load on the g_c factor instead loaded on three factors: the SRA

Math Concepts and Computation subtests loaded on the mathematical ability factor, the PPVT and the Gates-MacGinitie Vocabulary and Comprehension subtests loaded on the language ability factor, and the DSU rhyming fluency and originality scores loaded on the creativity factor. Of the four variables chosen as operationalizations of Level I intelligence and predicted to load on a memory factor, two, the PA test and the WISC-R digit span subtest, did load on a short term rote memory factor closely corresponding to Jensen's Level I, but the WISC-R Coding B subtest loaded on the mnemonic facility factor and the uncategorized free recall test failed to load significantly on any of the six factors.

These results give little support to the first hypothesis. In a further attempt to find evidence to uphold the existence of the constructs, 9 of the original 15 variables were selected to be used in a second factor analysis. The rationale for this analysis was that the g_c factor predicted in Hypothesis I represented a significant departure from Jensen's original theory of intelligence (Jensen, 1969). In the original theory, Level II intelligence was not split into crystallized (g_c) and fluid (g_f), but rather was considered a unitary analytic ability. Six of the seven operationalizations of g_c were dropped for this second analysis. The PPVT raw score was kept because Jensen (1970a) and Rohwer (1970) used it in earlier research as a measure of Level II intelligence. For the purposes of this study the PPVT had been classified as an operationalization of g_c due to its reliance upon culturally transmitted information. Jensen (1970a) also recognized this quality of the test, but found evidence that the test taps Level II processes in that it correlates .72 ($N = 638$) with Raven's CPM. With the removal of the 6 other variables classified as measures of g_c, it was anticipated that two factors corresponding to Jensen's earlier theory would emerge.

The WISC-R Coding B subtest, PA test, uncategorized free recall and the WISC-R digit span subtest were expected to load on one factor corresponding to Level I intelligence, and Raven's CPM, the WISC-R similarities subtest, the WISC-R block design subtest, the categorized free recall $\underline{Z}$ index and the PPVT raw score were expected to load on a second factor corresponding to Level II intelligence.

The matrix of intercorrelations among the 9 variables to be included in the second factor analysis, after partialing out chronological age as before, was subjected to a principal components analysis. Factors corresponding to eigenvalues greater than one were then rotated to oblique simple structure using the promax procedure. Three factors emerged with eigenvalues greater than 1.00; these eigenvalues are 2.0117, 1.3787, and 1.1154. The three factors accounted for 50.1 percent of the total variance. Table 9 presents the oblique factor loadings. The intercorrelations among the three oblique factors are presented in Table 10. The intercorrelations among the three oblique factors are all less than .15 in absolute value, suggesting that the three factors are relatively independent of one another.

Significant factor loadings from the second factor analysis are summarized in Table 11. Five variables loaded significantly on the first factor: the WISC-R Coding B subtest, uncategorized free recall, Raven's CPM, the WISC-R block design subtest and the categorized free recall $\underline{Z}$ index. This factor clearly does not correspond to either Level I or Level II intelligence as anticipated. Rather, 2 variables considered to be operationalizations of Level I and 3 variables considered to be operationalizations of Level II loaded on the factor. These 5 variables include all of those which loaded significantly on either the abstract reasoning using figural content factor or the mnemonic facility factor in the 15-variable solution, and 1 additional variable, the uncategorized free recall test, which previously had not loaded significantly on any factor. As pointed out in the presentation of the first factor analysis, the significant loadings of the WISC-R Coding B subtest, the categorized free recall $\underline{Z}$ index, Raven's

TABLE 9

FACTOR STRUCTURE OF THE ROTATED PRINCIPAL COMPONENTS OF THE NINE VARIABLES

Variable	Component I	Component II	Component III
WISC-R Coding B	.655*	-.312	.214
Paired-Associate Task	.156	-.103	.662*
Uncategorized Free Recall	.650*	-.008	-.102
WISC-R Digit Span	-.060	.100	.740*
Raven's CPM	.588*	.244	.112
WISC-R Similarities	.101	.729*	.096
WISC-R Block Design	.528*	.233	-.009
Categorized Free Recall Z Index	.486*	-.084	-.329
PPVT Raw Score	-.087	.823*	-.044

*Factor loadings significant at the .01 level (Harmon, 1967).

TABLE 10

INTERCORRELATIONS AMONG THE THREE OBLIQUE FACTORS

Factor	I	II	III
I	1.000		
II	-.149	1.000	
III	-.054	-.097	1.000

TABLE 11

SIGNIFICANT LOADINGS ON THE ROTATED FACTORS FROM THE ANALYSIS OF THE NINE VARIABLES

Variable	Factor		
	Strategy Usage	Verbal Intelligence	Short-Term Rote Memory
WISC-R Coding B	*		
Paired-Associate Task			*
Uncategorized Free Recail	*		
WISC-R Digit Span			*
Raven's CPM	*		
WISC-R Similarities		*	
WISC-R Block Design	*		
Categorized Free Recall Z Index	*		
PPVT Raw Score		*	

CPM and the WISC-R block design subtest on this factor suggest some intersection of mnemonic facility and abstract reasoning. The attribute common to both these factors which emerged in the preceding factor analysis is the necessity of discovering strategies for successful performance. These strategies may be rules or procedures which result in the successful completion of one of the matrices or the reproduction of one of the block designs, or they may be organizational structures used to facilitate the recall of arbitrary symbols or objects. In the light of these considerations, the factor might best be named strategy usage.

The significant loading of the uncategorized free recall test on this factor remains to be considered. There is extensive literature to suggest that on tasks where stimuli are not categorized a priori by the experimenter, subjects impose their own subjective organization (Tulving, 1962; Seibel, 1964; Bousfield, Puff, & Cowan, 1964). Tulving's index of subjective organization (<u>SO</u>) is an index developed for measuring the subject's tendency to "recall items in the same order on different trials in the absence of any experimentally manipulated sequential organization among items in the stimulus list [Tulving, 1962, p. 352]." Although the use of this index would have been an ideal method for determining the amount of strategizing or organizing children did on this task, results of the pilot test indicated that the index is unreliable when only four trials are administered. Therefore, it was not used as a variable in this study. However, it is well documented in the literature that "subjects do impose a sequential structure on their recall, that this subjective organization increases with repeated exposures and recall of the material, and that there is a positive correlation between organization and performance [Tulving, 1962, p. 352]." Such organization represents a type of strategy. In the light of this information the uncategorized free recall test does lend support to the interpretation of the first factor as representing strategy usage. The variables which loaded significantly on the second factor were the WISC-R similarities subtest and the PPVT raw score. This factor is very similar to the language ability factor found in the previous factor analysis except that the two reading achievement tests are not included. It is possible that this factor reflects verbal intelligence, since both the WISC-R similarities subtest and the PPVT were constructed to assess verbal intelligence; it will therefore be referred to as a verbal intelligence factor. The third factor in this solution corresponds perfectly to the short-term rote memory factor found in the first factor analysis. WISC-R digit span and the PA test load significantly on the factor. Hence, it will once again be referred to as the short-term rote memory factor, a close approximation to Jensen's Level I factor.

In summary, the second factor analysis is somewhat similar to the first six-factor solution. The strategy usage factor appears to be a combination of the mnemonic facility and abstract reasoning using figural content factors found in the first factor analysis. Both factor analyses have factors representing language or verbal abilities as well as a short-term rote memory factor. The results of the second factor analysis still do not provide support for the constructs of Level I and Level II intelligence. The first factor named, strategy usage, has tests purporting to measure both Level I intelligence and Level II intelligence loading on it. One implication of this finding may be that there are very few valid tests of Level I ability; one may reason that children will use strategies and organize stimuli input on almost any test, such that the test is no longer measuring Level I intelligence (rote memory) but rather the child's ability to strategize or organize. The verbal intelligence factor was not expected in that both the WISC-R similarities subtest and the PPVT were expected to load on the same factor as Raven's CPM, WISC-R block design, etc. if a unitary Level II factor existed. The short-term rote memory factor is in accordance with Level I intelligence, but as mentioned earlier 2 of the 4 variables expected to load on the factor loaded on the Strategy Usage factor instead. Inasmuch as the results of the two factor analyses failed to substantiate the existence of the constructs of Level I and Level II intelligence in either their original formulation or their reformulation using Cattell's constructs of fluid and crystallized intelligence (Jensen, 1973), Hypothesis I was not confirmed.

HYPOTHESIS II

Hypothesis II states that using two multitrait-multimethod matrices as described by Campbell and Fiske (1959), evidence of the convergent validity of each of the constructs of Level I and Level II intelligence will be provided as follows: the intercorrelations among the operationalizations of each of the two constructs in each matrix will be significantly greater than zero.

Building a multitrait-multimethod matrix requires two kinds of information: an intercorrelation matrix of all the variables involved and the test-retest or internal consistency reliabilities for each of the variables. When reliabilities are not available, they may be omitted (Campbell & Fiske, 1959). FACTOR3, a computer program maintained by the University of Wisconsin Academic Computing Center, was used to generate the matrix of intercorrelations among all 15 variables. Table 12 presents the values, type and sources of the reliability coefficients used in constructing the multitrait-multimethod matrices.

In presenting the multitrait-multimethod matrices a set of abbreviated labels will be used for the variables. Table 13 presents the abbreviations employed. Two matrices were constructed using all 15 variables to assess the convergent and discriminant validity of the constructs of Level I and Level II intelligence in the light of each of two possible sources of confounding method variance. For the first matrix the tasks were categorized by response mode--verbal versus nonverbal (see Table 3). Table 14 presents the first multitrait-multimethod matrix. Following the conventions employed by Campbell and Fiske (1959), various parts of the matrix are referred to by special names. Intercorrelations among operationalizations of the same construct which employ different methods are underlined. The underlined portions of the matrix are referred to as the validity diagonals, although they do not always appear as diagonals in the matrix. High values in the validity diagonals represent agreement between attempts to measure the same trait through different measures, providing evidence of convergent validity. Values in parentheses along the diagonal of the matrix are the reliability coefficients. The inter-correlations among the operationalizations of different traits which employ the same method are referred to as heterotrait-monomethod blocks, and are enclosed by a solid line. The significance of these values will be dealt with under Hypothesis III, which tests the discriminant validity of the constructs. Intercorrelations between variables having neither trait nor method in common are enclosed by broken lines and are said to occur in heterotrait-heteromethod blocks. The interpretation of these values is also reserved for Hypothesis III.

Hypothesis II may now be restated as follows: the values in the validity diagonals will all be significantly greater than zero as determined by the use of Fisher's _r_ to _Z_ transformation. Of the 3 values in the validity diagonal for Level I, all are positive and one is significant at the .01 level. Of the 18 values for Level II, all are positive, 12 are significant at the .05 level, and of these 9 are significant at the .01 level. In summary, one third of the values in the validity diagonal for the Level I construct and two thirds of the

TABLE 12

TYPE AND SOURCE OF RELIABILITIES INCLUDED ON THE MAIN DIAGONAL OF THE MULTITRAIT-MULTIMETHOD MATRICES

Variable	Value	Type	Source
WISC-R Coding B	.77	Test-Retest	Wechsler, 1974
Paired-Associate Task	.54	Test-Retest	Pilot Study
Uncategorized Free Recall	.41	Test-Retest	Pilot Study
WISC-R Digit Span	a		
Raven's CPM	.80	Test-Retest	Raven, 1965
WISC-R Similarities	.81	Test-Retest	Wechsler, 1974
WISC-R Block Design	.85	Test-Retest	Wechsler, 1974
Categorized Free Recall *Z* Index	.74	Test-Retest	Pilot Study
DSU Rhyming Fluency	.59	Test-Retest	Pilot Study
DSU Rhyming Originality	.63	Test-Retest	Pilot Study
PPVT Raw Score	.77	Test-Retest Using Parallel Forms	Dunn, 1965
SRA Math Concepts	.80	Internal Consistency (KR_{20})	Thorpe, Lefever, & Naslund, 1964
SRA Math Computation	.87	Internal Consistency (KR_{20})	Thorpe, Lefever, & Naslund, 1964
Gates-MacGinitie Vocabulary	.85	Test-Retest Using Parallel Forms	Gates & MacGinitie, 1965b
Gates-MacGinitie Comprehension	.83	Test-Retest Using Parallel Forms	Gates & MacGinitie, 1965b

[a]No reliability is available for the forward portion of the WISC-R digit span subtest. Test-retest reliability of the entire subtest is reported by Wechsler (1974) as .74.

TABLE 13

ABBREVIATIONS OF NAMES OF VARIABLES USED IN
MULTITRAIT-MULTIMETHOD MATRICES

Variable	Abbreviation
Operationalizations of Level I Intelligence	
WISC-R Coding B	Coding
Paired-Associate Task	PA
Uncategorized Free Recall	FRU
WISC-R Digit Span	DSF
Operationalizations of Level II Intelligence	
Raven's CPM	CPM
WISC-R Similarities	Sim
WISC-R Block Design	BD
Categorized Free Recall *Z* Index	FRC
DSU Rhyming Fluency	DSU-F
DSU Rhyming Originality	DSU-O
PPVT Raw Score	PPVT
SRA Math Concepts	SRA-Cn
SRA Math Computation	SRA-Cm
Gates-MacGinitie Vocabulary	GM-Voc
Gates-MacGinitie Comprehension	GM-Cmp

TABLE 14

MULTITRAIT-MULTIMETHOD MATRIX OF ALL FIFTEEN VARIABLES USING VERBAL VERSUS NONVERBAL RESPONSE MODE

	Verbal												Nonverbal		
	Level I			Level II									Level I	Level II	
	PA	FRU	DSF	Sim	FRC	DSU-F	DSU-O	PPVT	SRA-Cn	SRA-Cm	GM-Voc	GM-Cmp	Coding	CPM	BD
Verbal															
Level I															
PA	(.54)														
FRU	.03	(.41)													
DSF	.13	.01	( )												
Level II															
Sim	.12	.13	.11	(.81)											
FRC	.02	.15	-.08	.02	(.74)										
DSU-F	.04	.08	.11	.17	.04	(.59)									
DSU-O	-.01	.00	.01	-.09	.01	.26	(.63)								
PPVT	.00	.03	.12	.41	.03	.11	.04	(.77)							

TABLE 14 (cont.)

	Verbal												Nonverbal		
	Level I			Level II									Level I	Level II	
	PA	FRU	DSF	Sim	FRC	DSU-F	DSU-O	PPVT	SRA-Cn	SRA-Cm	GM-Voc	GM-Cmp	Coding	CPM	BD
SRA-Cn	-.00	.27	.15	.26	.08	.26	.06	.30	(.80)						
SRA-Cm	.16	.25	.06	.15	.05	.19	.01	.11	.65	(.87)					
GM-Voc	.25	.18	.31	.36	.06	.29	.08	.46	.52	.35	(.85)				
GM-Cmp	.13	.16	.22	.28	.14	.33	.04	.43	.49	.40	.76	(.83)			
Nonverbal															
Level I															
Coding	.12	.26**	.08	.04	.17	.22	-.06	-.07	.12	.20	.13	.19	(.77)		
Level II															
CPM	.09	.25	.13	.23*	.12	.22**	.04	.11	.37**	.30**	.26**	.24**	.18	(.80)	
BD	.09	.18	-.02	.14*	.04	.15*	.08	.13*	.18**	.19**	.07	.20**	.18	.38	(.85)

*Validity diagonal entry significant at the .05 level.

**Validity diagonal entry significant at the .01 level.

values in the validity diagonal for the Level II construct with respect to verbal versus nonverbal response mode are significant at the .05 level. These results provide some evidence of the convergent validity of the construct of Level II intelligence, but not of the construct of Level I intelligence.

A second matrix was constructed to assess the validity of the constructs of Level I and Level II intelligence in the light of a second possible source of confounding method variance. The 15 variables were categorized by administration type, either group or individually, as indicated in Table 3. The multitrait-multimethod matrix for this second categorization is presented in Table 15.

Of the three entries in the validity diagonal for Level I, all were positive but only one was significant. Of the 28 values in the validity diagonal for Level II, all but one were positive, 15 were significant at the .05 level, and of these 13 were significant at the .01 level. Once more, some evidence to support the convergent validity of Level II was provided, but little evidence to support the convergent validity of Level I was obtained.

Since it was thought appropriate to factor analyze only the nine "purer" measures of Level I and Level II intelligence, a second set of multitrait-multimethod matrices was constructed using only these nine "purer" variables and the same method categorizations as before. The nine-variable matrix used to assess the convergent and discriminant validity of the constructs of Level I and Level II intelligence with respect to verbal versus nonverbal response mode is presented in Table 16. Although the information presented in this table is included in Table 13, Table 16 was constructed to clarify the description of validity when using the nine "purer" variables.

The smaller multitrait-multimethod matrix employing response mode provides little support for the convergent validity of either Level I or Level II. Of the three validity diagonal entries for Level I, all are positive, but only one is significantly greater than zero. Of the six entries in the Level II validity diagonal all are positive, three are significantly greater than zero at the .05 level, and two are significant at the .01 level. In summary, no more than half of the entries in either validity diagonal differ significantly from zero.

Table 17 presents the fourth and last multitrait-multimethod matrix, employing the group versus individual administration categorization and only the nine "purer" variables (see Table 3). This last matrix also fails to provide strong evidence of the convergent validity of either Level I or Level II. Of the three validity diagonal entries for Level I, all are positive, but only one is significant. Of the four entries for Level II, all are positive but only two are significant.

After examination of the four multitrait-multimethod matrices, it is concluded that Hypothesis II concerning the convergent validity of the constructs is not supported with respect to either Level I or Level II intelligence. Even though substantial numbers of the correlations in the validity diagonals are statistically significant, the more important issue of psychological significance is not substantiated. The range of correlations in the validity diagonals of the four matrices for Level I is from .08 to .26, and the range of values for Level II is from .01 to .46. If correlations no greater than these can be obtained using independent methods to measure the same construct, the existence of the construct as a coherent psychological entity is clearly called into question.

In spite of the generally low intercorrelations among variables, the patterns evident in the correlation matrix appear reasonable. Correlations are generally positive, achievement tests are highly intercorrelated, and the

TABLE 15

MULTITRAIT-MULTIMETHOD MATRIX OF ALL FIFTEEN VARIABLES USING TYPE OF ADMINISTRATION

	Group Administration								Individual Administration						
	Level I	Level II							Level I			Level II			
	Coding	CPM	DSU-F	DSU-O	SRA-Cn	SRA-Cm	GM-Voc	GM-Cmp	PA	FRU	DSF	Sim	BD	FRC	PPVT
Group															
Level I															
Coding	(.77)														
Level II															
CPM	.18	(.80)													
DSU-F	.22	.22	(.59)												
DSU-O	-.06	.04	.26	(.63)											
SRA-Cn	.12	.37	.26	.06	(.80)										
SRA-Cm	.20	.30	.19	.01	.65	(.87)									
GM-Voc	.13	.26	.29	.08	.52	.35	(.85)								
GM-Cmp	.19	.24	.33	.04	.49	.40	.76	(.83)							

TABLE 15 (cont.)

	Group Administration								Individual Administration						
	Level I	Level II							Level I			Level II			
	Coding	CPM	DSU-F	DSU-O	SRA-Cn	SRA-Cm	GM-Voc	GM-Cmp	PA	FRU	DSF	Sim	BD	FRC	PPVT
Individual															
Level I															
PA	.12	.09	.04	-.01	-.00	.16	.25	.13	(.54)						
FRU	.26**	.25	.09	.00	.27	.25	.18	.16	.03	(.41)					
DSF	.08	.13	.11	.01	.15	.06	.31	.22	.13	.01	( )				
Level II															
Sim	.04	.23**	.17**	-.09	.26**	.15**	.36**	.28**	.12	.13	.11	(.81)			
BD	.18	.38**	.15*	.08	.18**	.19**	.07	.20**	.09	.18	-.02	.14	(.85)		
FRC	.17	.12	.04	.01	.08	.05	.06	.14*	.02	.15	-.08	.02	.04	(.74)	
PPVT	-.07	.11	.11	.04	.30**	.11	.46**	.43**	.00	.03	.12	.41	.13	-.03	(.77)

* Validity diagonal entry significant at the .05 level.

**Validity diagonal entry significant at the .01 level.

TABLE 16

MULTITRAIT-MULTIMETHOD MATRIX OF NINE VARIABLES
USING VERBAL VERSUS NONVERBAL RESPONSE MODE

	Verbal						Nonverbal		
	Level I			Level II			Level I	Level II	
	PA	FRU	DSF	Sim	FRC	PPVT	Coding	CPM	BD
Verbal									
Level I									
PA	(.54)								
FRU	.03	(.41)							
DSF	.13	.01	( )						
Level II									
Sim	.12	.13	.11	(.81)					
FRC	.02	.15	-.08	.02	(.74)				
PPVT	.00	.03	.12	.41	.03	(.77)			
Nonverbal									
Level I									
Coding	.12	.26**	.08	.04	.17	-.07	(.77)		
Level II									
CPM	.09	.25	.13	.23**	.12	.11	.18	(.80)	
BD	.09	.18	-.02	.14*	.04	.13*	.18	.38	(.85)

* Validity diagonal entry significant at the .05 level.

** Validity diagonal entry significant at the .01 level.

TABLE 17

MULTITRAIT-MULTIMETHOD MATRIX OF NINE VARIABLES USING TYPE OF ADMINISTRATION

	Group Administration		Individual Administration						
	Level I	Level II	Level I			Level II			
	Coding	CPM	PA	FRU	DSF	Sim	BD	FRC	PPVT
Group									
Level I									
Coding	(.77)								
Level II									
CPM	.18	(.80)							
Individual									
Level I									
PA	.12	.09	(.54)						
FRU	.26**	.25	.03	(.41)					
DSF	.08	.13	.13	.01	( )				
Level II									
Sim	.04	.23**	.12	.13	.11	(.81)			
BD	.18	.38**	.09	.18	-.02	.14	(.85)		
FRC	.17	.12	.02	.15	-.08	.02	.04	(.74)	
PPVT	-.07	.11	.00	.03	.12	.41	.13	.03	(.77)

* Validity diagonal entry significant at the .05 level.

** Validity diagonal entry significant at the .01 level.

correlations of the various WISC-R subtests with the CPM, the PPVT and the four achievement measures are reasonable, given that an individual subtest of the WISC-R would not be expected to correlate as highly as full scale WISC-R IQ with these measures. It is expected that if all the WISC-R subtests used in this research were combined, the resultant measure would correlate more highly with other IQ and achievement tests than did any individual subtest. These observations support the conclusion that the patterns evident in the correlation matrix are reasonable, even though the magnitudes of the correlations are generally low.

It must be noted that some of the reliabilities presented in the matrices and in Table 12 are low, and that none were obtained on the same sample which provided the intercorrelations. Hence, if problems occurred with these test administrations, or if the variables employed were in fact unreliable, an accurate assessment of the validity of the constructs may not be possible. There are several reasons for not having computed reliabilities of the variables using the same sample which yielded the intercorrelations. First of all, data on achievement test scores obtained from children's school records included only total raw scores. It was impossible to obtain item response data. Second, tests such as the DSU rhyming task and the WISC-R Coding B subtest are speeded, making internal consistency measures inappropriate. In these instances, pilot tests were conducted employing appropriate samples to obtain test-retest reliabilities which were then inserted in the matrices. Third, several more of the measures such as the categorized and uncategorized free recall tasks do not have separate items per se, precluding the computation of internal consistency reliabilities. In these instances, also, pilot tests using children of an appropriate age were conducted to establish test-retest reliabilities for use in the multitrait-multimethod matrices. The fourth reason for using reliability information from other sources was that of administrations requiring the establishment of a basal and a ceiling level, as is done with the PPVT and all of the WISC-R subtests employed except Coding B. Typically, no subject may receive all of the items on a test of this type, again necessitating the use of a test-retest reliability in this context. In the light of these considerations, the conclusion drawn concerning the convergent validity of the constructs must be interpreted with caution.

HYPOTHESIS III

Hypothesis III states that by using two multitrait-multimethod matrices as described by Campbell and Fiske (1959), evidence of the discriminant validity of each of the constructs of Level I and Level II intelligence will be provided as follows: in each matrix, the correlation between any two operationalizations of the same construct will be higher than either of their correlations with any operationalization of the other construct. To test this hypothesis, values on the validity diagonals are compared first with values in the same row or column enclosed in the heterotrait-heteromethod blocks (indicated by broken lines) and then with the values in the same row or column enclosed in the heterotrait-monomethod blocks (indicated by solid lines). Values in the validity diagonals which are higher than values in the same row or column enclosed by broken lines (heterotrait-heteromethod block) indicate that a verbal response mode test and a nonverbal response mode test of the same construct correlate more highly than a verbal response mode test and a non-verbal response mode test of different constructs. Values in the validity diagonals which are higher than values in the same row or column enclosed by solid lines (monotrait-heteromethod block) indicate that a verbal response mode test and a nonverbal response mode test of the same construct correlate more highly than a test of Level I and a test of Level II which employ the same response mode. A less stringent test of discriminant validity, also implied by Hypothesis III, is that the values in the monotrait-monomethod triangles (neither enclosed nor underlined) also be greater than values lying in the same row or column and enclosed by either broken or solid lines. This means that two measures of the same construct employing the same response mode should correlate more highly with each other than with measures of the other construct, regardless of response mode employed.

For the matrix employing verbal versus nonverbal response mode and all 15 variables, presented in Table 14, 304 comparisons of values are implied. For 70 of these comparisons it is predicted that two measures of Level I will correlate more highly with each other than with measures of Level II. Of these 70 comparisons less than half, only 31, are in the predicted direction. For the remaining 234 comparisons it is predicted that two measures of Level II will correlate more highly with each other than with measures of Level I. Of these 234 comparisons just under two-thirds, 154, are in the predicted direction. In summary, tests of Level I tend to correlate more highly with tests of Level II than with each other, and tests of Level II also tend to correlate more highly with other tests of Level II than with tests of Level I. This finding could be explained in part by the lower reliabilities of the Level I tests. Results of these 304 comparisons fail to support the discriminant validity of the constructs.

Upon inspection of the matrix, no salient method effect appears either, indicating that performance on neither Level I nor Level II tests is affected by response mode (verbal versus nonverbal). A method effect would be indicated by higher values in the heterotrait-heteromethod blocks, where the median correlation was .13, and by higher values in the monotrait-monomethod triangles, where the median correlation was .16, than in the validity diagonals, where the median correlation was .18.

It might be argued that an effect of response mode exists only within Level I or within Level II. In looking for a method effect within Level I, the intercorrelations among tests of Level I employing the same response mode are compared with those employing different response modes. The median correlation within the Level I monomethod blocks is .03 and the median correlation within the Level I heteromethod blocks is .12. This suggests that there is no effect of response mode within Level I. Within Level II the median monomethod correlation is .19, and the median heteromethod correlation is .17, again providing little evidence that type of response mode (verbal versus nonverbal) affects performance.

At this point results concerning the discriminant validity of the constructs as tested by the second 15 variable multitrait-multimethod matrix will be presented. The matrix being examined (see Table 15) employs a method classification of group versus individual administration. For this matrix 391 comparisons of values are implied. For 87 of these comparisons it is predicted that two measures of Level I will correlate more highly with each other than with measures of Level II. Of these 87 comparisons only 40, less than half, are in the predicted direction. For the remaining 304 comparisons it is predicted that two measures of Level II will correlate more highly with each other than with measures of Level I. Of these 304 comparisons 215, or 71 percent are in the predicted direction.

In summary, tests of Level II are once again seen to correlate more highly with all the other variables than tests of Level I do. Again this result could be due to the higher reliabilities of the Level II tests. Although intercorrelations among tests of Level II are higher than intercorrelations between Level I and Level II tests in 71 percent of the cases, it should be noted that the magnitudes of most correlations in the matrix are so low that evidence of discriminant validity would be weak even if all comparisons were in the predicted direction.

There appears to be no salient method effect, indicating that type of administration failed to affect performance on either Level I or Level II tests. A method effect would be indicated by higher values in the heterotrait-monomethod blocks, where the median correlation was .12, than in the heterotrait-heteromethod blocks, where the median correlation was .17. Another indication of a method effect would appear if there were higher values in the monotrait-monomethod triangles than in the validity diagonals. The median correlation for values in the monotrait-monomethod triangles was .23, as compared to the median correlation of .14 in the validity diagonals. As indicated, there is little support for a method effect. Looking more closely at the effect of type of administration within Level I only, the median correlation within the Level I heteromethod blocks is .12, while the median correlation within the Level I monomethod triangle is only .03. If the median correlation within the Level I tests using the same type of administration were higher than the median correlation of Level I tests using different types of administration, then there would be evidence of a method effect. As the results indicate there is no support for a method effect within the Level I tests. The same sets of correlations are compared for the Level II tests. The median correlation within the Level II monomethod triangle is .26 while the median correlation within the Level II heteromethod blocks is .13, giving some support for a method effect. Once again results must be regarded cautiously, given the low magnitudes of correlations involved.

The next few paragraphs will review the results concerning the discriminant validity of the first of the nine-variable multitrait-multimethod matrices. In constructing the first of these smaller matrices the tasks were classified according to response mode (verbal versus nonverbal) as indicated in Table 3. The matrix is presented in Table 16. For this nine-variable matrix 85 comparisons are implied. For 39 of these comparisons it is predicted that two measures of Level I will correlate more highly with each other than with measures of Level II. Of these 39 comparisons 25, or 64 percent, are in the predicted direction. For the remaining 46 comparisons it is predicted that two measures of Level II will correlate more highly with each other than with measures of Level I. Of these 46 comparisons just over half, or 24, are in the predicted direction. In summary, evidence of discriminant validity provided by the third matrix is very weak. Once more the magnitudes of the correlations reported are too low to support any strong conclusions. However, the uniformity of the correlations across the various parts of the matrix argues against the existence of the constructs.

The search for a method effect as it was carried out for the first two matrices would be redundant, as only the set of Level II tests has been changed. Accordingly, inspection of the matrix with regard to the effect of response mode was curtailed to only an examination of the Level II tests. Within the Level II monomethod blocks the median correlation was .21, while the median correlation within the Level II heteromethod block was .13, providing minimal support for an effect of response mode within the "purer" measures of Level II.

The last matrix, presented in Table 17, is used to assess the discriminant validity of the nine "purer" measures employing the method classification of type of administration (group versus individual). For this matrix 85 comparisons of values are implied. For 33 of these comparisons it is predicted that two measures of Level I will correlate more highly with each other than with measures of Level II. Of these 33 comparisons only 17, or just over half, are in the predicted direction. For the remaining 52 comparisons it is predicted that the two measures of Level II will correlate more highly with each other than with measures of Level I. Of these 52 comparisons 35, or 66 percent, are in the predicted direction. There is virtually no support for the discriminant validity of the constructs given these results.

An examination of any possible effect of type of administration (group versus individual) within the set of "pure" measures of Level II intelligence employed in the fourth matrix gave no evidence of any method effect. Within Level II the median monomethod correlation was .09, while the median correlation across methods was .18.

In summary, the discriminant validity of the Level I and Level II constructs is not supported by the four multitrait-multimethod matrices, and the results predicted in Hypothesis III are not obtained. Whether the lack of confirmation is attributable to the nonexistence of coherent psychological traits or due to the inclusion of improper operationalizations of Level I and Level II is unclear. One conclusion that may be drawn from inspection of each of the matrices is that the correlations between measures of Level I and measures of Level II tend to be higher than the intercorrelations among measures within Level I. This suggests that the operationalizations of Level I chosen for use in this study may not be performing as expected. These results may be explained in part by differences in the types of performance elicited in each Level I test. If the items or tasks presented in the various measures of Level I intelligence are amenable to solution by cognitive operations or processes other than rote learning, then some children may have employed some cognitive processes in solving Level I tests which are shared with the Level II tests.

SUPPLEMENTARY ANALYSIS

The results presented in the previous sections focused on the existence of the constructs. This section will focus on the relationship of each of the four measures of achievement to each of the other tests administered.

Table 18 presents the Pearson product-moment correlations of each of the four achievement measures with each of the remaining 11 variables. Thirty of the 44 correlations are significant at an alpha level of $p < .05$, and of these 26 are significant at an alpha level of $p < .01$. These correlations were computed on the entire sample of 221 children.

The high number of significant correlations is not surprising, and supports the contention that the study of Level I and Level II intelligence as operationalized by the variables included in this study is of relevance to the school environment. Although the magnitudes of the correlations presented in Table 18 are not high, some positive relationship is evident between scholastic achievement measures and each of the other variables, with the exception of the DSU originality score.

TABLE 18

PEARSON PRODUCT-MOMENT CORRELATIONS OF ACHIEVEMENT MEASURES WITH EACH OF THE REMAINING VARIABLES

	Achievement Measures			
Variable	SRA Math Concepts	SRA Math Computation	Gates-MacGinitie Vocabulary	Gates-MacGinitie Comprehension
WISC-R Coding B	-.120	.202**	.127	.191**
Paired Associate Task	-.004	.155*	.249**	.131
Uncategorized Free Recall	.270**	.245**	.179**	.157*
WISC-R Digit Span	.150*	.055	.305**	.220**
Raven's CPM	.365**	.295**	.256**	.236**
WISC-R Similarities	.264**	.148*	.357**	.278**
WISC-R Block Design	.181**	.191**	.072	.197**
Categorized Free Recall Z Index	.084	.046	.059	.142*
DSU Rhyming Fluency	.255**	.194**	.287**	.328**
DSU Rhyming Originality	.063	.007	.080	.038
PPVT Raw Score	.302**	.105	.459**	.425**

* $p < .05$

** $p < .01$.

CONCLUSIONS

Tests of Hypotheses I, II, and III have failed to support the existence, convergent validity, or discriminant validity of the constructs. Therefore, three alternative conclusions must be considered. In Chapter V, each of the following alternatives will be discussed: (1) The operationalizations may have been inappropriate, (2) the theory from which the constructs arose may be inaccurate, or (3) some aspect of the experimental procedures or methodology may have precluded finding the expected results.

V

DISCUSSION

Three possible explanations will be considered for the lack of support for the hypotheses as stated. First, the adequacy of the operationalizations will be discussed. Second, methodological weaknesses will be examined. Third, the results concerning each of Hypotheses I, II, and III will be considered in relation to the constructs of Level I and Level II intelligence and alternative theoretical frameworks. Finally, educational implications will be presented.

ADEQUACY OF OPERATIONALIZATIONS AND METHODOLOGY

Discussion of the adequacy of the operationalizations and the methodological weaknesses of the study might typically be placed in a section entitled "critique and limitations." The present organization was chosen to clarify the interpretation of the study as a construct validation, where failures of operationalizations or experimental procedures may in themselves be of consequence in evaluating the validity of the constructs.

Adequacy of Operationalizations

The definitions of both Level I and Level II intelligence are in terms of internal processes. Those internal processes must be logically inferred on the basis of the characteristics of the measures used, e.g., when only one trial on the PA task is administered, it is assumed that the amount of mediation possible is quite limited, or when obvious categories are provided in the stimulus materials for a free recall task, there is ample opportunity for the application of mnemonic strategies to facilitate recall. While it is clear that some transformation or manipulation of the input is required to solve the Raven's CPM, for example, it is difficult to guarantee that the internal processes elicited in each subject by an intended operationalization of Level I intelligence will not involve any transformation or manipulation of the input. Given that so much of Jensen's definitions is concerned with the specification of internal processes and so little discussion of observable behaviors is provided, it is difficult to assure the adequacy of the operationalizations employed. As indicated in Chapter IV, there is evidence that some of the measures intended as operationalizations of Level I intelligence elicited Level II processes in at least some of the children tested.

Adequacy of Experimental Procedures and Methodc .ogy

Several methodological weaknesses may have contributed to the failure to confirm Hypothesis I. One problem was that of low reliabilities on several of the instruments used. Such reliabilities preclude high intercorrelations among the various operationalizations of each construct, affecting both the factor analytic solutions and the multitrait-multimethod matrices

A second problem was that the correlations of the PA test and Raven's CPM with the other variables may have been attenuated due to the levels of difficulty of these two tasks. The mean and standard deviation of the PA test indicated little variation in performance, and some ceiling effect was evident in the CPM. Again, these problems may have reduced the probability of finding support for the constructs.

A third factor which may have resulted in somewhat lower intercorrelations than otherwise was the wide range of testing conditions encountered in the various schools. It was not always possible to secure facilities for quiet and uninterrupted individual test administrations. While data collected under adverse conditions may have attenuated the obtained correlations, testers were carefully trained in advance and supervised during the testing to ensure the greatest possible uniformity in test administrations.

ADEQUACY OF THE THEORIES IN WHICH THE CONSTRUCTS OF LEVEL I AND LEVEL II INTELLIGENCE ARE EMBEDDED

This paper is oriented toward the substance and methodology most familiar to the individual differences psychologist. However, some recent findings by cognitive and developmental psychologists also provide support for the interpretations presented.

Hypothesis I

In the light of the problems of operationalization and methodology discussed above, inferences from this research must be made with caution. Nonetheless, the possibility remains that the failure to confirm Hypothesis I implies that Level I and Level II are not distinct abilities.

Six-Factor Solution. If Level I and Level II intelligence indeed represent two independent types of intelligence, then factor analysis of a suitable battery of tests should yield factors corresponding to these two constructs. All of the tests in the battery, with the possible exception of the DSU rhyming task which loaded on the creativity factor, should, according to Jensen's (1969) theory be composed of some combination of Level I and Level II intelligence. The relative proportions of Level I and Level II intelligence required by the various tests should vary, even though no test would be expected to be a pure measure of either one. The creativity test might be expected to be different from the others, in that although creativity is related to intelligence it clearly involves other factors as well (Cronbach, 1970).

It was predicted that, aside from the creativity test scores, three factors would emerge. Not only were five other factors found, but four anomalies distinguished the obtained factor structure from the predicted structure. First, all operationalizations of Level I intelligence did not load on the same factor. Second, all operationalizations of crystallized Level II intelligence did not load on the same factor. Third, all operationalizations of fluid Level II intelligence did not load on the same factor.

Fourth, some operationalizations of Level I intelligence and some operationalizations of Level II fluid loaded significantly on the same factor. Subject to the limitations of this study, these anomalies require either the introduction of new constructs or modifications of existing ones. The incorporation of still more constructs, or types of intelligence, to account for the additional factors obtained is not acceptable. Cronbach and Meehl (1955) caution that any such modification (e.g., splitting a concept into two or more portions) must be tested on a different body of data prior to acceptance, to avoid the danger of a posteriori rationalization. Furthermore, the introduction of additional constructs is not sufficient to explain the results of the factor analysis of all 15 variables. Mathematical ability, language ability, abstract reasoning using figural content, and creativity might all be interpreted as types of Level II intelligence, but the variables loading significantly on the mnemonic facility factor include proposed operationalizations of each of Level I and Level II, precluding the interpretation of the factor as a type of either Level I or Level II intelligence. In the light of these results, Level I intelligence, Level II fluid, and Level II crystallized intelligence may not be valid constructs.

Since the results of the first factor analysis do not correspond to the prediction in Hypothesis I, alternative explanations for the obtained factor structure will be considered next.

It might be expected that the factor structure obtained in the first analysis would correspond to the predictions of one or more of the multiple factor theories of intelligence reviewed in Chapter II. No unitary factor appeared which could be interpreted as corresponding to Spearman's "G", and, while Cattell's two-factor theory might account for the six factors, explanations in terms of specific content, etc. would have to be sought to explain the fracturing of each of his factors of fluid and crystallized mental abilities. Five of the six factors obtained in the present study might be interpreted as representing five of Thurstone's (1938) primary mental abilities. The factor labeled abstract reasoning using figural content resembles Thurstone's spatial ability (S), short term rote memory corresponds to Thurstone's memory (M), mathematical ability corresponds to number (N), language ability corresponds to verbal (V), and creativity resembles word fluency (W). The mnemonic facility factor, however, does not correspond to either of Thurstone's two remaining abilities, reasoning (R) and perceptual speed (P). While Thurstone's theory provides a partial interpretation of the data, most of the six factors obtained are so test-specific that the evidence provided in support of any theoretical interpretation is weak. Only two tests loaded significantly on five of the six factors. To assess the adequacy of Thurstone's theory to explain the results obtained, additional research employing many more measures would have to be conducted. None of the later intelligence theories, including Guilford's (1967) structure of intellect model, provide any better explanation of the six factors.

The mnemonic facility factor is explained by none of the theories reviewed. The variables loading significantly on mnemonic facility include intended operationalizations of each of Level I and Level II intelligence. This factor foreshadows a problem with the concept of a distinct memory factor: it is difficult to maintain the construct of Level I intelligence when operationalizations of Level I intelligence load on factors presumably representing types of Level II intelligence.

The mnemonic facility factor is distinctive in several respects. No analagous factor appears in the work of any of the major intelligence theorists, and the argument presented in Chapter IV for the labeling of the factor employed terms and drew upon literature not typically encountered in individual differences psychology. The free recall paradigm is not usually employed in individual differences psychology, and the two free recall tests in the present study both load more highly on mnemonic facility than on any other factor, suggesting a possible explanation for the failure of intelligence theorists to have described this factor.

The generation or modification of mnemonic strategies is an example of a more general feature of human thought processes. Given a novel problem-solving situation, an individual will invent or adapt idiosyncratic strategies to facilitate performance (Hagan, 1972). Flavell (1970) describes such strategies to facilitate performance as "mnemonic mediation." Such mediation is characterized as "a planful, instrumental, cognitive act, akin to problem-solving behavior (Flavell, 1970, p. 208)." The retrieval of information may be regarded as a particular type of problem to be solved, and mnemonics are a particular class of strategies which facilitate the solution of the problem of remembering. In addition to the free recall tests and the WISC-R Coding B subtest, all of which loaded on the mnemonic facility factor, four other measures included in the battery reflect a child's performance in novel problem-solving situations where skills or rules required for solution are not taught in school. These are the PA test, WISC-R digit span, WISC-R block design and Raven's CPM. The PA test and WISC-R digit span were designed to discourage strategizing, as discussed in Chapters III and IV. Miller, Galanter, and Pribram (1960, p. 136) state that "in tests of immediate memory . . . subjects seldom try any mnemonic tricks--with only one presentation of the material, there is little time to develop a Plan [for remembering] and little need of it." The number of items to be recalled in each of these tests, up to eight pairs of pictures or up to nine digits, are within the "7 $\pm$ 2" units suggested by Miller (1956) as the limit to the capacity of immediate memory, which also suggests that mnemonic strategies should not be required for successful performance on these two tests. The remaining two novel problem-solving situations, Raven's CPM and the WISC-R block design subtest, have been described as requiring complex reasoning strategies (Wechsler, 1949; Glasser & Zimmerman, 1967; Raven, 1965). The perceptual factors involved in the performance of these tests as well as the different nature of the strategies required for successful performance distinguish these tests from the tests which loaded on the mnemonic facility factor, and may explain the appearance of the factor labeled abstract reasoning using figural content. The skills required for successful performance on the DSU Rhyming test, WISC-R similarities, PPVT, and the four tests of scholastic achievement are much more likely to have been taught in school, and these tests are not characterized as novel problem-solving situations.

Three-Factor Solution. Since the first factor analysis failed to support Hypothesis I as stated, an additional analysis not implied in Hypothesis I was performed in an attempt to find support for the existence of Level I and Level II intelligence. The rationale, description, and discussion of the second factor analysis are provided below.

In consideration of the possible inadequacies of the operationalizations employed in this study, and in the light of the ad hoc nature of Jensen's (1973) employment of Cattell's constructs, it was decided to attempt a validation of the constructs of Level I and Level II intelligence as originally proposed, using only measures which clearly corresponded to Jensen's (1969) definitions of the constructs and/or which had been employed by various researchers. Nine variables were included in the second factor analysis. Using these nine "purer" measures, a three-factor solution was obtained, which failed to validate the two-factor theory of intelligence (see Chapter IV).

The three factors obtained in the second analysis, described in Chapter IV, were labeled strategy usage, verbal intelligence, and short term rote memory. The short term rote memory factor was also found in the first analysis. The verbal intelligence factor corresponds exactly to the language ability factor found in the first analysis, except that two of the four variables with significant loadings on language ability were not included in the second factor analysis. The strategy usage factor in the second analysis includes exactly those variables which loaded on the mnemonic facility factor and on the factor labeled abstract reasoning using figural content. The uncategorized free recall test, which loaded marginally on the mnemonic facility factor in the first analysis, loaded significantly on strategy usage in the second analysis. Both of the variables which loaded on the short term rote memory factor were intended as operationalizations of Level I intelligence, and both of the variables loading on verbal intelligence were intended as operationalizations of Level II intelligence. However, two of the four operationalizations of Level I intelligence and three of the five operationalizations of Level II intelligence loaded significantly on the factor labeled strategy usage. These included the uncategorized free recall test, intended as an operationalization of Level I, as well as the categorized free recall test and the CPM, both chosen to operationalize Level II. Thus, the strategy usage factor does not correspond to either Level I or Level II intelligence.

Because the obtained factor structure failed to support the existence of the constructs of Level I and Level II intelligence, alternative explanations of the results of the second factor analysis will be considered.

None of the major intelligence theorists describe a factor, such as strategy usage, which saturates tests of both complex reasoning and memory. The factor cannot be equated to Spearman's "G", because it is distinct from verbal intelligence as measured by the WISC-R similarities subtest and the PPVT. Strategy usage may appear to correspond to Thurstone's reasoning (R) factor, but that factor is characterized as involving logical reasoning, which does not appear to be required for performance on the free recall and coding tests. The failure of intelligence theorists to report such a factor may be due in part to the factor's reliance upon tests more frequently used in experimental psychology than in the study of individual differences. Even Cronbach's (1970) spectrum for comparing tests of scholastic aptitude or general ability fails to accomodate such experimental tasks as the PA test or the two free recall tests. This is significant in that Cronbach's spectrum is atheoretical and locates a wide variety of tests along a continuum ranging from those requiring maximum training to those requiring maximum adaptation or transfer. It therefore gives little guidance for interpreting the strategy usage factor.

Given that intelligence theories have not frequently dealt with experimental tasks, it is useful to turn toward the body of psychological literature which contends that heuristics or mnemonics are critical processes necessary for organizing and retrieving information. Some of this literature was discussed with reference to the mnemonic facility factor which emerged in the six-factor solution.

A posteriori, all of the five tests which loaded on the strategy usage factor may logically be shown to be similar in an essential way. Each requires the acquisition of heuristics, mnemonics or tricks to make the learning task easier. Each requires the formulation or acquisition of a rule, procedure, or method of organizing information to facilitate performance, rather than the application of previously acquired information.

It is reasonable to suppose that this requirement for the adaptation or generation of novel strategies in approaching these five tests might account for the emergence of the strategy usage factor. Although not all researchers have employed the terms "strategy" or "heuristic" in their work, many processes facilitative to problem solving may be regarded as examples of the employment of strategies. These processes include the use of mnemonic mediators (Flavell, 1970); verbal rehearsal of strings of words (Flavell, 1970); verbal mediation hypothesis (Spiker, Gerjuoy, & Shepard, 1956; Stevenson & McBee, 1958; Reese, 1961); types of strategies documented in probability learning (Stevenson & Weir, 1963); selective attention for separating task-relevant and task-irrelevant aspects of a situation (Hagan, 1972); coding, labeling, and categorizing (Maccoby & Hagen, 1965); hypothesis generation and testing (Bruner, Goodnow, & Austin, 1956); and win-stay, lose-shift strategies found in the study of learning sets (Berman, Rane, & Bahow, 1970). These various types of mediators, strategies, heuristics, etc. may all be described as involving "planfulness," i.e. engaging in some activity in the anticipation of improving performance (Flavell, 1970). Hagan (1972) alludes to this generalization when he states that rehearsal strategies will be implemented only when the subject has a plan indicating that it will be useful to do so. The importance of plans, strategies, heuristics, etc. is emphasized by Miller, Galanter and Pribrum (1960) in their book <u>Plans and the Structure of Behavior</u>.

In summary, it is reasonable to suppose that children more facile in developing novel approaches to new problems will do better on the WISC-R Coding B subtest, categorized and uncategorized free recall tests, the WISC-R block design subtest, and Raven's CPM.

Given that plans and strategies are central to performance on such a wide variety of psychological tasks, why didn't all of the tests load on the strategy usage factor? As discussed with reference to the six-factor solution, the two tests which loaded on the short term rote memory factor were specifically designed to discourage strategizing, and the immediate recall of short lists of stimuli should not require recoding for successful performance. The two tests which loaded on the verbal intelligence factor, the WISC-R similarities subtest and the PPVT, may require strategies for successful performance, but these tests are much less novel in nature and the requisite strategies are much more likely to have been taught in school. Thus, performance on these tests may be less dependent upon the child's facility in developing strategies in new situations.

<u>Hypothesis II</u>

The discussion of results concerning Hypothesis II will be brief. In testing this hypothesis it is possible to determine the appropriateness of the constructs of Level I and Level II intelligence by examining the agreement among operationalizations of each construct which are specified in advance. This approach is in opposition to that of the factor analytic technique, where any result is potentially possible and the psychologist must attempt to make sense of whatever factors emerge.

The two matrices constructed using all 15 variables and employing verbal versus nonverbal response mode and group versus individual administration as method classifications gave marginal support to the convergent validity of Level II, but little support to the convergent validity of Level I. The two matrices employing only the nine "purer" measures and the same method classifications failed to substantiate the convergent validity of either Level I or Level II intelligence.

If in fact the various tests included in the battery were completely unrelated, some low negative and some low positive correlations would have been expected. As indicated in Chapter IV, the correlations between variables representing attempts to measure the same construct by different methods ranged from .08 to .26 across the four matrices for Level I and from .01 to .46 across the four matrices for Level II. The fact that all of the values in the validity diagonals of the four matrices were low and positive supports the contention that the various operationalizations of each construct do have some variance in common, but not enough to be of great psychological interest.

In the light of the factor analytic results discussed with reference to Hypothesis I, the findings concerning the convergent validity of the constructs are not surprising. Little evidence for the convergent validity of Level I was obtained, suggesting that the construct as defined and operationalized is not a coherent psychological entity.

The matrices employing all 15 variables gave more support to the convergent validity of Level II than the matrices employing only 9 variables, indicating that the evidence for the convergent validity of Level II in the larger matrices depended primarily upon the 6 variables which were dropped to form the set of nine "purer" measures. These 6 variables were least defensible as operationalizations of Level II intelligence.

Hypothesis III

The discussion of the results concerning Hypothesis III will also be brief. Campbell and Fiske (1959) pointed out that the convergence of indicators, i.e., positive correlations among different types of measures, is insufficient to substantiate a construct. Some distinction between a given construct and other constructs, as evidenced by low correlations between measures of the different constructs, is also required. Cronbach (1972) derives the same requirement from considerations of parsimony. Different scientific names should not be applied to the same entity or to two constructs very similar in what they measure, thus needlessly complicating the theory; hence the requirement of discriminant validity.

The four multitrait-multimethod matrices constructed in this study gave essentially the same results concerning the discriminant validity of the constructs. Little support was provided for the hypothesis that Level I and Level II are distinct, and in general operationalizations of Level I tended to correlate more highly with Level II operationalizations than with each other. Comparisons among the various parts of the matrices also failed to reveal any strong method effects, although some evidence was found of an effect of group versus individual administration among the Level II tests. Once more, the lack of support is not surprising in the light of the factor analytic results discussed with reference to Hypothesis I.

Each of the methods used in forming a multitrait-multimethod matrix serves as a counter-hypothesis which may account for observed differences in behavior

(Webb, Campbell, Schwartz, & Sechrest, 1966). The particular methods examined in this study did not prove to be significant determinants of behavior. Stevenson et al. (1971) pointed the way toward the examination of methods used in this study. Although this research attempted to overcome some of the weaknesses in the Stevenson et al. (1971) research, limitations in the methodology employed in this study preclude drawing definitive conclusions.

IMPLICATIONS OF THE RESEARCH

Implications must be drawn from these research findings cautiously, given the methodological weaknesses of the study. With so little support for the existence of the proposed constructs being found, applications in the field of education cannot be seriously considered. The short term rote memory factor which emerged in both factor analyses with only the PA test and the digit span test loading on it is interpreted as a highly artifactual ability which has virtually no application in the educational environment. It seems that most information which is presented in the school situation permits organization through any number of strategies or heuristics available to the child. If this argument is valid then of what use is the existence of the Level I construct, given that it may neither predict nor explain any behavior of educational consequence?

The question of the validity of Level II must also be considered. When the 15 variable factor analysis was performed two of the factors which emerged were mathematical ability and language ability. The mathematical ability factor had only two tests loading significantly on it: SRA Math Concepts and SRA Math Computation. The language ability factor had four tests loading on it, two of which were subtests of the Gates-MacGinitie Reading Test. When the 15 variable multitrait-multimethod matrices were examined there was some marginal support for the existence of Level II. These results are of interest to the educator, as achievement tests have been used for diagnostic and selection purposes (Cronbach, 1972). However, when these less "pure" measures are removed for the second factor analysis and the 9 variable multitrait-multimethod matrices are constructed, support for the construct of Level II decreases. What remains in terms of the second factor analysis is the strategy usage factor and the verbal intelligence factor. An interesting question for the educational psychologist is what does the strategy usage factor allow us to predict or explain in the school situation that could not be predicted by using current intelligence tests? Does Level II have much relevance to the school environment? Since Raven's CPM, figure copying, and categorized free recall are the measures of Level II most frequently used in researching the construct, while such conventional measures as WISC-R subtests and the Stanford-Binet have not been used, the utility of Level II in predicting scholastic achievement remains to be established.

It has been shown that performance on tests which are highly analytic in nature, such as the WISC-R block design (Vigotsky, 1962), and tests of conceptual maturity such as conservation of weight (Smedslund, 1961), may be facilitated by direct training. However, Smedslund's (1961) experiment suggests that the gains effected by such training may be lost when the testing paradigm is slightly altered. Cronbach (1972) summarizes research studies of this type by saying that responses required by any set of test items can indeed be taught, but educators cannot be interested in training for test tasks alone. The aim of educational intervention is to develop a child's ability to perform on a wide range of intellectual tasks. A successful method of training on Raven's CPM, for example, would only contribute to intellectual development if it improved a child's performance on other unpracticed tasks as well. Cronbach (1972) claims that currently we have no clearly valid methods for developing analytic abilities faster than they could emerge in a normally rich environment. Given that Level II intelligence was formulated as a highly abstract, analytic ability, its relevance to educational intervention is unclear.

REFERENCES

Academic Computing Center. FACTORS: Factor analysis. Computer Services-Statistics Center, Madison: University of Wisconsin, 1974.

American Psychological Association, Committee on Ethical Standards in Psychological Research. Ethical principles in the conduct of research with human participants. Washington, D.C., 1973.

Berman, P.W., Rane, N.G., & Bahow, E. Age changes in children's learning set with win-stay, lose-shift problems. Developmental Psychology, 1970, 2, 233-239.

Bialer, I. Primary and secondary stimulus generalization as related to intelligence level. Journal of Experimental Psychology, 1968, 62, 395-402.

Binet, A., & Henri, V. La psychologie individuelle. L'Annee psychologique 2, 1895. Translated by M.D. Boring. In R.J. Herrnstein & E.G. Boring (Eds.), A source book in the history of psychology. Cambridge, Mass.: Harvard University Press, 1966. Pp. 428-433.

Binet, A., & Simon T. Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. Année Psychologie, 1905, 11, 191-244. In J.C. Nunnally (Ed.), Psychometric theory. New York: McGraw Hill, 1967.

Boring, E.G. A history of experimental psychology. (2nd ed.) New York: Appleton-Century-Crofts, 1950.

Bousfield, W.A. The occurrence of clustering in the recall of randomly arranged associates. Journal of Genetic Psychology, 1953, 49, 229-240.

Bousfield, W.A., & Cohen, B.H. The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates. Journal of Psychology, 1953, 36, 67-81.

Bousfield, W.A., & Cohen, B.H. The occurrence of clustering in the recall of randomly arranged words of different frequencies-of-usage. Journal of Genetic Psychology, 1955, 52, 83-95.

Bousfield, W.A., Puff, C.R., & Cowan, T.M. The development of constancies in sequential organization during repeated free recall. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 489-495.

Bruner, J.S., Goodnow, J.J., & Austin, G.A. A study of thinking. New York: Wiley, 1956.

Burt, C. The structure of the mind: A review of the results of factor analysis. British Journal of Educational Psychology, 1949, 19, 100-111.

Campbell, D., & Fiske, D. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Cattell, R.B. Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 1963, 54, 1-22.

Cole, M., Frankel, F., & Sharp, D. Development of free recall learning in children. Developmental Psychology, 1971, 4, 109-123.

Cronbach, L.J. Essentials of psychological testing. (3rd ed.) New York: Harper, 1970.

Cronbach, L.J. Test validation. In R.L. Thorndike (Ed.), Educational measurement. (2nd ed.) Washington, D.C.: American Council on Education, 1972. Pp. 443-507.

Cronbach, L., & Meehl, P. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

Dunn, L.M. Expanded manual: Peabody Picture Vocabulary Test. Minneapolis: American Guidance Service, 1965.

Flavell, J.H. The developmental psychology of Jean Piaget. Princeton, N.J.: Van Hostrand, 1963.

Flavell, J.H. Developmental studies of mediated memory. In H. Reese & L. Lipsitt (Eds.), Advances in child development and behavior (Vol. 5). New York: Academic Press, 1970. Pp. 181-211.

Frankel, F., & Cole, M. Measures of category clustering in free recall. Psychological Bulletin, 1971, 76, 39-44.

Friedrichs, A.G., Hertz, T.W., Moynahan, E.D., Simpson, W.E., Arnold, M.R., Christy, M.D., Cooper, C.R., & Stevenson, H.W. Interrelations among learning and performance tasks at the preschool level. Developmental Psychology, 1971, 4, 164-172.

Garrett, H.E. A developmental theory of intelligence. American Psychologist, 1946, 1, 372-378.

Gates, A.I., & MacGinitie, W.H. Teacher's Manual, Gates-MacGinitie Reading Tests. New York: Teacher's College Press, 1965. (a)

Gates, A.I., & MacGinitie, W.H. Technical manual, Gates-MacGinitie Reading Tests, survey D. New York: Teacher's College Press, 1965. (b)

Glasser, A., & Zimmerman, I. Clinical interpretation of the Wechsler Intelligence Scale for Children. New York: Greene & Stratton, 1967.

Guilford, J.P. Creativity. American Psychologist, 1950, 5, 444-454.

Guilford, J.P. The structure of intellect. Psychological Bulletin, 1956, 53, 267-293.

Guilford, J.P. The nature of human intelligence. New York: McGraw Hill, 1967.

Hagan, J.W. Strategies for remembering. In S. Farnham-Diggory (Ed.), Information processing in children. New York: Academic Press, 1972.

Harmon, H H. Modern factor analysis. (2nd ed.) Chicago: University of Chicago Press, 1967.

Hendrickson, A.E., & White, P.O. Promax: A quick method for rotation to oblique simple structure. British Journal of Statistical Psychology, 1964, 17, 65-70.

Inhelder, B., & Piaget, J. The early growth of logic in the child. Translated by E.A. Lunzer & D. Papert. New York: Harper & Row, 1964.

Jensen, A.R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Jensen, A.R. Comparison of "culture-loaded" and "culture-fair" tests. In A.R. Jensen & W.D. Rohwer, Jr. (Eds.), An experimental analysis of learning abilities in culturally disadvantaged children. Final Report, U.S. Office of Economic Opportunity Contract 2404. Washington, D.C.: U.S. Office of Education, July 1970. (a)

Jensen, A.R. The organization of abilities in preschool children. In A.R. Jensen & W.D. Rohwer, Jr. (Eds.), An experimental analysis of learning abilities in culturally disadvantaged children. Final Report, U.S. Office of Economic Opportunity Contract 2404. Washington, D.C.: U.S. Office of Education, July 1970. (b)

Jensen, A.R. A two level theory of mental abilities. In A.R. Jensen & W.D. Rohwer, Jr. (Eds.), An experimental analysis of learning abilities in culturally disadvantaged children. Final Report, U.S. Office of Economic Opportunity Contract 2404. Washington, D.C.: U.S. Office of Education, July 1970. (c)

Jensen, A.R. Level I and Level II abilities in three ethnic groups. American Educational Research Journal, 1973, 10, 263-276.

Jensen, A.R., & Frederiksen, J. Free recall of categorized and uncategorized lists: A test of the Jensen hypothesis. Journal of Educational Psychology, 1973, 65, 304-312.

Kohs, S.C. Intelligence measurement: A psychological and statistical study based upon the Block-Design Tests. New York: Macmillan, 1923.

Maccoby, E.E., & Hagen, J. Effects of distraction upon central versus incidental recall: Developmental trends. Journal of Experimental Child Psychology, 1965, 2, 280-289.

Mandler, G. Organization and memory. In K. Spence & J.T. Spence (Eds.), The psychology of learning and motivation (Vol. 1). New York: Academic Press, 1967.

Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Miller, G.A., Galanter, E., & Pribram, K.H. Plans and the structure of behavior. New York: Holt, 1960.

Peterson, J. Early conceptions and tests of intelligence. New York: Harcourt, Brace & World, 1925.

Raven, J.C. Guide to using the Coloured Progressive Matrices: Sets A, AB, B. London: H.K. Lewis, 1965.

Reese, H.W. Transposition in the intermediate size problem by preschool children. Child Development, 1961, 32, 311-314.

Rohwer, W.D., Jr. Mental elaboration and learning proficiency. In A.R. Jensen & W.D. Rohwer, Jr. (Eds.), An experimental analysis of learning abilities in culturally disadvantaged children. Final Report, U.S. Office of Economic Opportunity Contract 2404. Washington, D.C.: U.S. Office of Education, 1970.

Rohwer, W.D., Jr., Ammon, M.S., Suzuki, N., & Levin, J.R. Ethnicity--SES and learning proficiency. In A.R. Jensen & W.D. Rohwer, Jr. (Eds.), An experimental analysis of learning abilities in culturally disadvantaged children. Final Report, U.S. Office of Economic Opportunity Contract 2404. Washington, D.C.: U.S. Office of Education, July 1970.

Seibel, R. An experimental paradigm for studying the organization and strategies utilized by individual Ss in human learning and an experimental evaluation of it. Paper presented at meetings of The Psychonomic Society, Niagara Falls, Ontario, October 1964.

Smedslund, J. The acquisition of conservation of substance and weight in children. Scandinavian Journal of Psychology, 1961, 2, 1-10; 71-84; 85-87; 153-160; 203-210.

Spearman, C. The abilities of man. New York: Macmillan, 1927.

Spiker, C.C., Gerjuoy, I.R., & Shepard, W.O. Children's concept of middle-sizedness and performance on the intermediate size problem. Journal of Comparative and Physiological Psychology, 1956, 49, 416-419.

Stevenson, H.W., & McBee, G. The learning of object and pattern discriminations by children. Journal of Comparative and Physiological Psychology, 1958, 51, 752-754.

Stevenson. H.W., & Weir, M.W. The role of age and verbalization in probability learning. American Journal of Psychology, 1963, 76, 299-305.

Stevenson, H.W., Williams, A.M., & Coleman, E. Interrelations among learning and performance tasks in disadvantaged children. Journal of Educational Psychology, 1971, 62, 179-184.

Thorndike, E.L., Bregman, E.O., Cobb, M.V., & Woodyard, E. The measurement of intelligence. New York: Teacher's College, Columbia University, 1927.

Thorpe, L.P., Lefever, D.W., & Naslund, R.A. SRA Achievement Series Test coordinator's manual for forms C and D. Chicago: Science Research Associates, 1964.

Thurstone, L.L. Vectors of mind. Chicago: University of Chicago Press, 1935.

Thurstone, L.L. Primary mental abilities. Psychometric Monographys, 1938, 1, 11-121.

Tulving, E. Subjective organization in free recall of "unrelated" words. Psychological Review, 1962, 69, 344-354.

Underwood, B.J. Psychological research. New York: Appleton-Century-Crofts, 1957.

Vernon, P.E. The structure of human abilities. New York: Wiley, 1950.

Vigotsky, L.S. Thought and language. New York: Wiley, 1962.

Webb, E.J., Campbell, D.T., Schwartz, R.D., & Sechrest, L. Unobtrusive measures: Nonreactive research in the social sciences. Chicago: Rand McNally, 1966.

Wechsler, D. Manual for the Wechsler Intelligence Scale for Children. New York: Psychological Corporation, 1949.

Wechsler, D. Manual for the Wechsler Intelligence Scale for Children--Revised. New York: Psychological Corporation, 1974.

White, S.H. Evidence for a hierarchical arrangement of learning processes. In L.P. Lipsitt & C.C. Spiker (Eds.), Advances in child development and behavior. Vol. 2. New York: Academic Press, 1965.

APPENDIX A

SPECIFIC INSTRUCTIONS AND MATERIALS USED IN THE ADMINISTRATION OF UNPUBLISHED TESTS

In the following sections, verbatim instructions prepared for use in this study are presented. For the PA task, pictorial stimuli are also included.

WISC-R CODING B

Look at these divided boxes or squares (point to key). You see each box has a number in the top part and a special mark in the bottom part. Each number has its own different mark. Now look down here where it says SAMPLE (point to sample), where the boxes have numbers in the top part, but the squares on the bottom are empty. You are to put in the empty squares the marks that should go there like this (demonstrate using blackboard): Here is a 2. The 2 has this mark (write in symbol on board). So I put it in this square like this. Here is a 1. The 1 has this mark (write symbol on board). So I put it in this square. Now you fill in all of the boxes up to the red line. Don't skip any! (Walk around the class to see if they understand.) When I tell you to start, you do the rest of them. Begin after the red line and fill in as many squares as you can, one after the other, without skipping any. Keep working until I tell you to stop. Work as quickly as you can without making mistakes. When you finish the first line, go on to the second. Go ahead. Do them in order. Don't skip any. (At the end of 120 seconds say) Stop, put your pencils down.

FREE RECALL (CATEGORIZED AND UNCATEGORIZED)

I am going to show you some objects. As I show you each object, tell me its name. Look at each object carefully and try to remember it. After I have shown you all the objects, I will ask you to tell me the names of as many of them as you can. Watch carefully. (Present each object for 2 seconds. With the presentation of the first few objects, say) What is this? (After all objects are presented, say) Now name as many of the objects as you can. (Allow 90 seconds for the child to respond. Then say) Good. Now see if you can do as well this time. (Move the boxes to second predetermined random order. Then say) Tell me the name of each object as I show it to you. (Present each object for 2 seconds. After all objects are presented, say) Name as many of the objects as you can. (Allow 90 seconds for the child to respond. After child finishes naming the objects, say) That was a good job. (As you move the boxes to the next predetermined order, say) I will show you the objects two more times. Try to do better each time. (Present each of the objects; have child name them. Then say) Go ahead. (Allow 90 seconds for the child to respond.) Fine. This is the last time. (Move the boxes to the last predetermined order. Then present each of the objects; have child name them. Then say) Go ahead. (Allow 90 seconds for child to respond.)

PA TASK

I am going to show you some pairs of pictures. Look at each pair carefully and try to remember which pictures go together. After I show you all the pairs, I will show you one picture from each pair and ask you what picture goes with it.

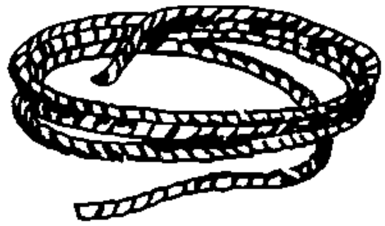
Let's try a few. Try to remember which pictures go together. (Present example pairing trials at 5 second presentation time. Wait 5 seconds, then present singleton pictures. As each singleton is presented say) Tell me what goes with this. (If child makes an error, correct him by showing the card with the correct pair, then continue. After practice test, say) Now we are ready to begin. Do you understand what you are to do? There will be more pairs this time, so look at each pair carefully. (Present pairs at 5 second presentation time. When first singleton of the test trial is presented, say) What goes with this? (Continue presenting singletons. Let child pace him/her self, allowing no more than a 10-second presentation for each singleton.)

DSU RHYMING TASK

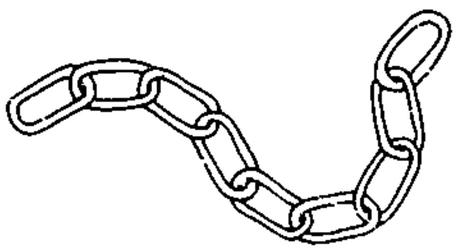
You will have one minute to write down all the words you can think of that rhyme with cat. Begin writing now. (Place placard bearing the word "cat" on chalk tray. At the end of one minute, say) Stop. Put your pencils down.

PICTORIAL STIMULI USED FOR PA TASK

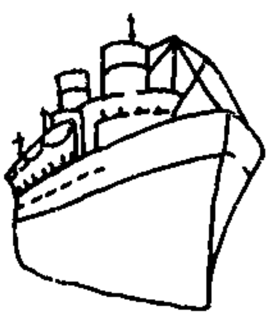
The pictures used for the practice test and for the 8-pair PA test are presented on the following pages. Pairs are shown as they were presented during the study trial. For the test trial, only the left picture from each pair was presented. Test trial presentations were in a different order than study trial presentations.



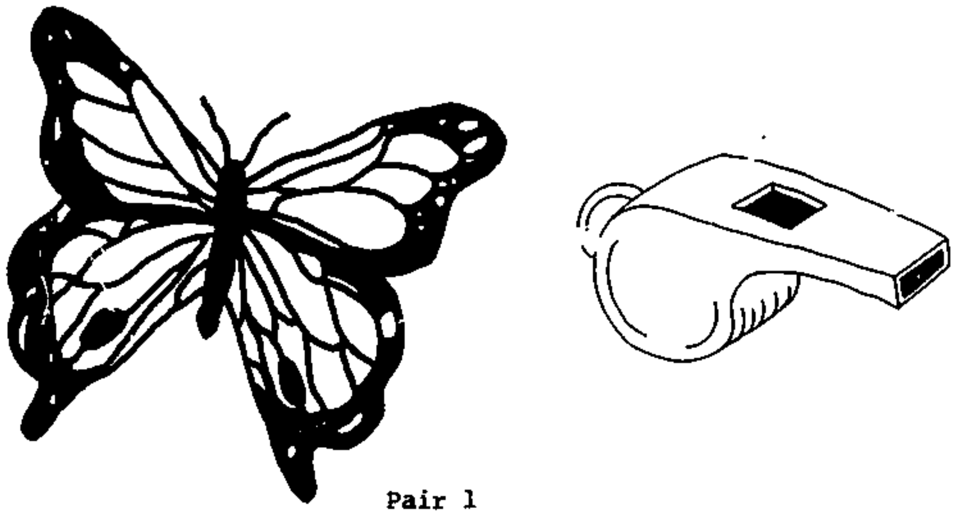
Sample pair 1



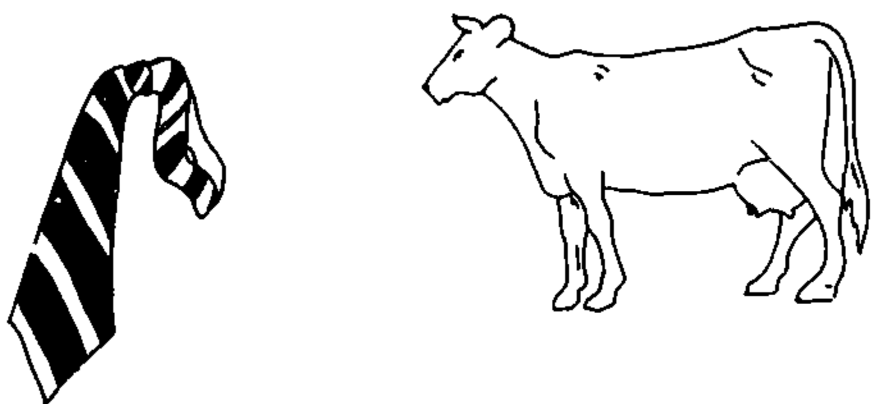
Sample pair 2



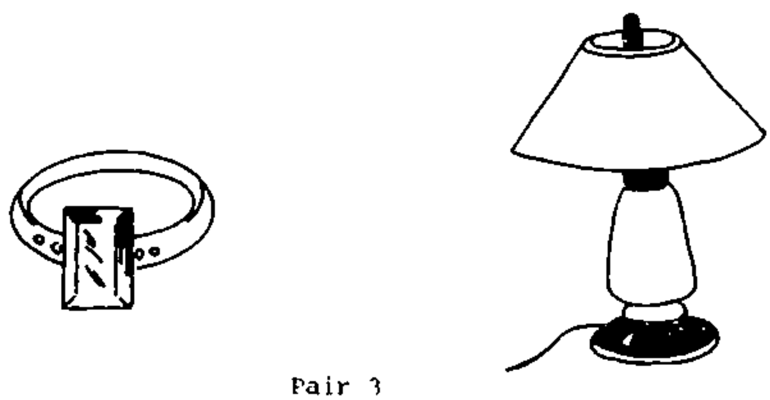
Sample pair 3



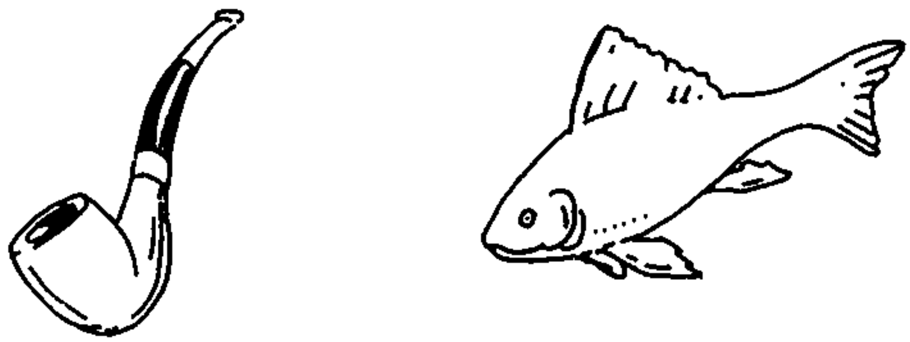
Pair 1



Pair 2

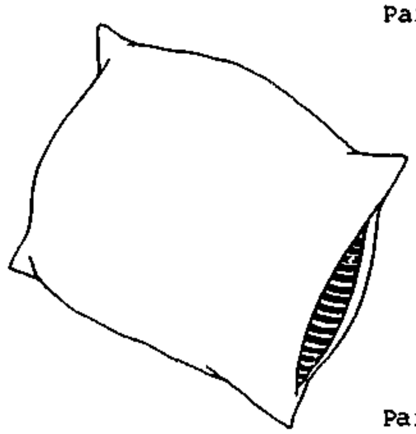


Pair 3



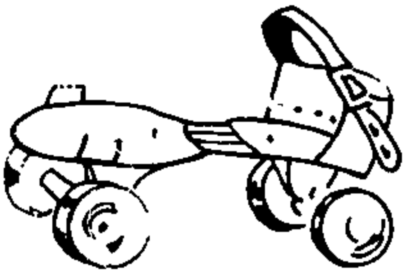
Pair 4

Pair 5



Pair 6

Pair 7



Pair 8

APPENDIX B

EXPLANATORY LETTER AND PARENTAL
PERMISSION FORM

October 14, 1974

To the Parents of ________________________________:

In cooperation with the [name of district] elementary schools, the Wisconsin Research and Development Center for Cognitive Learning, a department of the University of Wisconsin, will conduct research from October 7 to November 7, 1974, in each elementary school. Only students in fifth grade are to be included in the study.

The purpose of the research is to learn more about how children learn. The data gathered will shed some light on the acceptability of several recent theories about learning and intelligence. Each little bit that we can learn will help to improve the educational process in schools.

This letter is being sent to each fifth grader's parents to tell them about the study and to obtain written permission for their child to participate. Each child will be given a series of short memory and intelligence tests. These will be given during the regular school dav at the child's school. About ninety minutes of testing will be involved. Testing will be done in several short sessions, with some tests being given individually and others in classroom-sized groups.

All test administration will be done by trained personnel from the R&D Center and from the University of Wisconsin School of Education. Children usually find these tests to be interesting and even game-like, and respond positively during the testing.

As is customary with this type of testing, the test administrators will help to make children comfortable and will give encouragement during the testing, but they will not give the children an evaluation of their performance on the tests. Since the tests are experimental, we will not provide the schools or teachers with any information on individual children's performance, although we will inform them of group results. No test scores will be entered into your child's school records as a result of this study.

All data will be treated confidentially by the R&D Center. To assure anonymity, children's names will be removed from papers before the data is analyzed. No names will be used in the published reports nor will there be any way to identify an individual's performance. The findings will be expressed in terms of the relationships among various tests which emerge when large groups of scores are considered together.

We will also need information about each child's school achievement. According to Public Law 93-380 (the Education Amendments of 1974) and Wisconsin Statutes, parents of children participating in such studies are entitled to know specifically what information from the child's school records will be used in the study. For the purposes of this study, we would like your permission to obtain from your child's school records the Vocabulary and Comprehension subtest scores from the Gates-MacGinitie Reading Test and the Computation and Concepts subtest scores from the Math portion of the SRA Skills Profile, both administered during the first semester of the 1973-1974 school year.

The nature and design of the study make it necessary to have certain information about each child's family background. The four questions on the attached permission sheet indicate the needed information. You may be assured that this information will also be handled in the most confidential manner possible.

When you have completed and signed the form, place it in the envelope provided and have your child deliver it to his or her teacher. You may add comments on the margins or back of the form. Even if you should decide for some reason not to allow your child to participate, you should return the form to the school. This will assure that all have been contacted, and that no child will be left out inadvertently.

If you have a question or desire further information, contact Geneva DiLuzio, who has primary responsibility for this research project. She may be reached at the Research and Development Center, 262-4901.

Your cooperation and assistance are both important and appreciated. Every child who participates increases the reliability of the information gathered in this study, and the study is designed to improve education for all through a better understanding of learning processes.

Most sincerely,

Geneva DiLuzio
Evaluation Specialist

[name of superintendent]
Superintendent of Schools

To Geneva DiLuzio (Wisconsin Research and Development Center for Cognitive Learning) and to teachers at my/our child's school: Permission is hereby granted for my/our child, (child's first and last name) to be tested in the research study to be conducted by the R&D Center from approximately October 7 through November 7, 1974 in the [name of district] schools, as described in Geneva DiLuzio's letter of October 14, 1974. I grant permission for the release of the Vocabulary and Comprehension subtest scores from the Gates-MacGinitie Reading Test, and the Computation and Concepts subtest scores from the Math portion of the SRA Skills Profile, both administered during the first semester of the 1973-1974 school year to Geneva DiLuzio from existing school records for use in this study. I understand that the results of the testing will be used in confidence for research purposes only and will not become a part of my child's school records.

______________________________ ____________________
(signature of parent/guardian) (date)

I/We have answered the four questions below in order to give needed family information about the child.

1) How many younger brothers and sisters does the child have?

_________ none

_________ 1 or 2

_________ 3 or more

2) How many older brothers and sisters does the child have?

_________ none

_________ 1 or 2

_________ 3 or more

3) The parents' formal education is best described as: (Check one in each column)

Father	Mother	
____	____	No high school
____	____	Some high school
____	____	High school graduate
____	____	Some college
____	____	College graduate
____	____	Advanced degree (please specify what degree)

4) The present occupation of parents is: (Former occupation if retired)

Father ______________________________

Mother ______________________________

PLEASE RETURN BY: [date one week prior to beginning testing in school.]

[The actual permission slip and family background questions were typed on a single page. The two parts were cut apart, with only a number associating family background information with the remainder of each child's materials.]

National Evaluation Committee

Francis S. Chase, Chairman
Emeritus Professor
University of Chicago

Helen Bain
Past President
National Education Association

Lyle Bourne
Professor
University of Colorado

Sue Buel
Consultant, Portland, Oregon

Roald F. Campbell
Emeritus Professor
The Ohio State University

George E. Dickson
Dean, College of Education
University of Toledo

Larry R. Goulet
Professor
University of Illinois

Chester W. Harris
Professor
University of California - Santa Barbara

William G. Katzenmeyer
Professor
Duke University

Barbara Thompson
Superintendent of Public Instruction
State of Wisconsin

Joanna Williams
Professor
Teachers College
Columbia University

University Advisory Committee

John R. Palmer, Chairman
Dean
School of Education

William R. Bush
Deputy Director
R & D Center

David E. Cronon
Dean
College of Letters and Science

Diane H. Eich
Specialist
R & D Center

Evelyn L. Hoekenga
Coordinator
R & D Center

Dale D. Johnson
Associate Professor
Curriculum and Instruction

Herbert J. Klausmeier
Member of the Associated Faculty
R & D Center

James M. Lipham
Member of the Associated Faculty
R & D Center

Wayne R. Otto
Associate Director
R & D Center

Richard A. Rossmiller
Director
R & D Center

Elizabeth J. Simpson
Dean
School of Family Resources
and Consumer Sciences

Len Van Ess
Associate Vice Chancellor
University of Wisconsin - Madison

Associated Faculty

Vernon L. Allen
Professor
Psychology

B. Dean Bowles
Professor
Educational Administration

Thomas P. Carpenter
Assistant Professor
Curriculum and Instruction

Marvin J. Fruth
Professor
Educational Administration

John G. Harvey
Professor
Mathematics
Curriculum and Instruction

Frank H. Hooper
Professor
Child Development

Herbert J. Klausmeier
V.A.C. Henmon Professor
Educational Psychology

Joseph T. Lawton
Assistant Professor
Educational Psychology

Joel R. Levin
Professor
Educational Psychology

L. Joseph Lins
Professor
Institutional Studies

James M. Lipham
Professor
Educational Administration

Donald N. McIsaac
Professor
Educational Administration

Gerald Nadler
Professor
Industrial Engineering

Wayne R. Otto
Professor
Curriculum and Instruction

Robert G. Petzold
Professor
Music
Curriculum and Instruction

Thomas S. Popkewitz
Assistant Professor
Curriculum and Instruction

Thomas A. Romberg
Professor
Curriculum and Instruction

Richard A. Rossmiller
Professor
Educational Administration

Dennis W. Spuck
Assistant Professor
Educational Administration

Michael J. Subkoviak
Assistant Professor
Educational Psychology

Richard L. Venezky
Professor
Computer Sciences

J. Fred Weaver
Professor
Curriculum and Instruction

Larry M. Wilder
Assistant Professor
Child Development